

Virtual Counselor Extenders: Social Cue Taxonomy, Soft Rooms, and Brief Motivational Interviewing

CHRISTINE LISETTI and STEPHANIE LUNN, Florida International University, USA
JAMES HODGINS, DIANA MARIN, and JAYNISH HAZARI, Loma Linda University, USA
UBBO VISSER, University of Miami, USA
MAYA BOUSTANI, Loma Linda University, USA

To address behavioral health shortages, we propose virtual counselor extenders: socially interactive virtual health agents (VHAs) that operationalize evidence-based interventions to expand clinicians' reach and mitigate barriers to access for stigmatized populations. Current VHAs often lack systematic therapeutic communication and evidence-based environmental design. We present five contributions: (1) the implementation of a complete Brief Motivational Interviewing (BMI) intervention for heavy-drinking adults; (2) the application of a social cue taxonomy to systematically document multimodal active listening behaviors; (3) the design of a therapeutic "soft room" grounded in environmental psychology; (4) a quantitative evaluation of usability and engagement with the target population; and (5) a qualitative thematic analysis of user acceptance regarding agent demographics and environmental realism. We conclude by extending the taxonomy with therapeutic-specific cues and providing design recommendations for future virtual counseling systems. This work demonstrates that autonomous agents can deliver complex therapeutic interventions while maintaining high user acceptance.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **HCI design and evaluation methods**; **User studies**; • **Applied computing** → **Health Informatics**; **Consumer health**; **Health care information systems**;

Additional Key Words and Phrases: socially interactive agents, embodied conversational agents, virtual agents, conversational user interfaces, digital mental health interventions, motivational interviewing, alcohol use disorder

ACM Reference Format:

Christine Lisetti, Stephanie Lunn, James Hodgins, Diana Marin, Jaynish Hazari, Ubbo Visser, and Maya Boustani. 2018. Virtual Counselor Extenders: Social Cue Taxonomy, Soft Rooms, and Brief Motivational Interviewing. *ACM Trans. Comput.-Hum. Interact.* 37, 4, Article 111 (August 2018), 66 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 Introduction

The global growing demand for accessible behavioral and mental health services has driven significant research into virtual socially intelligent agents designed in the role of "counselor extenders" [24]: digital technology that can expand and complement clinicians' therapeutic support by delivering evidence-based standardized high-quality screening, evaluation, and brief treatments [82, 98, 198]. Globally, more than 1 billion people live with mental health conditions, but the median availability of mental-health professionals is only about 13 per

Authors' Contact Information: Christine Lisetti, lisetti@cs.fiu.edu; Stephanie Lunn, sjlunn@fiu.edu, Florida International University, Miami, Florida, USA; James Hodgins, JHodgins@students.llu.edu; Diana Marin, Jaynish Hazari, Loma Linda University, Los Angeles, USA; Ubbo Visser, University of Miami, Coral Gables, Florida, USA, visser@cs.miami.edu; Maya Boustani, Loma Linda University, Los Angeles, California, California, USA, mboustani@llu.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, or post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 1557-7325/2018/8-ART111

<https://doi.org/XXXXXXXX.XXXXXXX>

100,000 population [196]. In the United States alone, the U.S. National Center for Health Workforce Analysis indicates severe shortages across behavioral healthcare professions by 2037, with supply projected to only meet 57% of demand for mental health counselors, 55% for psychologists, 45% for addiction counselors, and about 43% for adult psychiatrists [65]. This need is particularly acute for stigmatized conditions, where individuals may be reluctant to seek traditional counseling [31, 104], creating opportunities for virtual agents to provide initial low-intensity interventions for stepped care when people are too self-conscious or anxious to engage with human counselors at first [30, 72, 112]. Unlike chatbots or Large Language Models (LLM) applications, virtual agents are 3-dimensional (3D) computer graphics characters that inhabit a 3D virtual environment, and run a loop of perception-cognition-action enabling them to perceive and express human verbal and nonverbal behaviors (e.g., speech, facial expressions, gestures) in realtime. With these abilities, virtual agents can address the main limitations identified by mental health professionals about chatbots – namely, their inability to communicate nonverbally [179]. However, the effectiveness of virtual health agents (VHAs) - virtual socially interactive agents designed with features specific to healthcare contexts [160?] - as counselor extenders depends on their ability to model believable therapeutic communicative behaviors, deliver evidence-based interventions, and operate within environments that support therapeutic engagement.

Active listening constitutes an essential therapeutic communication skill, which consists of coordinated verbal and nonverbal behaviors that demonstrate engagement, empathy, and understanding [155]. In human counseling relationships, therapists employ systematic embodied listening strategies that include strategic questioning combined with sustained eye contact, forward-leaning posture, responsive nodding, to create psychological comfort and encourage client disclosure [38, 76].

Current virtual counseling systems exhibit limitations in modeling systematic active listening behaviors. Existing virtual healthcare interviewer agents like SimSensei [39] demonstrate sophisticated capabilities but mainly focus on detecting and analyzing user behaviors, rather than generating coordinated therapeutic listening responses. While recent machine learning approaches [54, 92, 200] have achieved promising results in generating natural-looking gestures, these data-driven methods present challenges for therapeutic contexts where behavior appropriateness and clinical interpretability must be ensured. Furthermore, virtual agent research often treats asking, listening, and explaining as separate behavioral components, rather than as coordinated communicative acts within a unified therapeutic framework.

Additionally, virtual agent research has not systematically applied evidence-based therapeutic office environment design principles. Environmental factors including color schemes, natural elements, and spatial arrangements have been shown to affect psychological states and therapeutic outcomes [41, 184], yet many virtual counselor implementations operate in generic settings [138] that may not support therapeutic engagement. Importantly, a recent review revealed that researchers frequently did not include sufficient details on which theories they applied in their design of the health intervention content, or on the agent specific features to enable reproduction of results [89]. As clinicians extenders, however, VHAs should only deliver evidence-based health interventions to ensure the validity, reproducibility, efficacy, and safety of therapeutic material they provide to vulnerable populations [24].

This work addresses these limitations by presenting a virtual counseling system that integrates systematic active listening behaviors, evidence-based intervention delivery, and therapeutic environment design within a web-accessible platform. Our approach operationalizes established therapeutic communication principles by mapping specific nonverbal behaviors to each communicative act.

Our research questions (RQs) are the following:

- **RQ1:** *What design choices can inform the development of a virtual health agent as counselor that elicits similar socio-emotional responses as those experienced with an effective human counselor?*

- **RQ2:** *What design choices can inform the development of a virtual office environment for a virtual agent counselor that triggers similar socio-emotional responses as those experienced in an effective human counselor's physical office?*
- **RQ3:** *What socio-emotional responses do diverse adults exhibit when they can choose from a diverse set of virtual health agents as counselors for a counseling intervention?*

The primary contributions of this research are fivefold. First, we adapt and develop a complete Brief Motivational Interviewing intervention for virtual agent delivery, building on evidence-based Hester's Drinker's Check-Up [73] for web-based delivery to heavy-drinking adults recruited online. Second, we present a detailed taxonomic design framework that systematically integrates individual social cues, multimodal active listening behaviors, and therapeutic environment features using Feine et al.'s social cue taxonomy [51], in an effort to facilitate the accumulation of systematic knowledge on the multitude of cues involved in socially interactive agent design. Third, we present the design and implementation of a therapeutic "soft room" environment incorporating counseling literature principles including calming colors, natural elements, ample seating, among other features. Fourth, we discuss a quantitative evaluation by target population measuring the VHA's usability, usefulness, engagement, and human-likeness characteristics. Fifth, we provide a qualitative analysis through in-depth interviews with heavy-drinking adults, identifying key factors influencing user acceptance. Our findings reveal three primary themes: *agent acceptability* (personal qualities, appearance, nonverbal behaviors, voice characteristics), *agent diversity* (age, racial representation, gender considerations), and *agent environment* (therapeutic space design, realism, views). Based on these empirical findings, we extend an existing taxonomy of social cues with new cues specific to therapeutic communication contexts, and derive design recommendations for VHAs from our mixed methods approach.

The remainder of this paper is organized as follows.

- (1) Section 2 reviews existing approaches to multimodal behavior generation in virtual agents, and describes digital evidence-based brief motivational interventions;
- (2) Section 3 discusses how our virtual agent counselor system design is grounded in three theoretical frameworks: a user acceptance model; a social science taxonomy of virtual agent individual social cues used to document detailed design decisions for virtual agent features; and metaphorical design of virtual counselors based on social science counseling theories and principles;
- (3) Section 4 presents our implementation of an evidence-based Brief Motivational Interviewing (BMI) intervention content we adapted for virtual agent delivery;
- (4) Section 5 describes our comprehensive taxonomic design framework for virtual agent counselors, including individual social cues, combinations of individual social cues for systematic active listening behaviors, and therapeutic environment features;
- (5) Section 6 discusses the quantitative evaluation of the VHA from target population who completed the BMI intervention;
- (6) Section 7 explains the data gathering and analysis process that we conducted for the qualitative analysis of participants' interviews;
- (7) Section 8 presents evaluation results of user responses and thematic findings;
- (8) Section 9 discusses the implications of our findings and design recommendations for future virtual counseling systems;
- (9) Section 10 proposes an extended taxonomy of social cues for VHAs;
- (10) Section 11 identifies limitations with our study and implications;
- (11) Section 12 summarizes our contributions and concludes with future directions.

By demonstrating that virtual agents can deliver complete therapeutic interventions while maintaining user acceptance, this work provides a foundation for scalable mental health interventions that may reach underserved populations and supplement traditional therapeutic services.

2 Related Research

This section defines the concept of *virtual health agents* as counselor extenders, as it relates to other types of autonomous agents. It also discusses related research on virtual agents, and introduces brief motivational interviewing interventions (BMIs) as one of the most effective evidence-based behavior change intervention, pointing to limitations of text-only BMI digital health interventions.

2.1 Definition of Virtual Health Agents (VHA)

Our work introduces and operationalizes the concept of Virtual Health Agents (VHA), a specialized application of intelligent virtual agent (IVA) technology [97?] designed specifically for therapeutic communication and evidence-based intervention delivery. Drawing from the recent comprehensive taxonomy of socially interactive agents [105], we extend these autonomous capabilities to encompass clinical roles. As terminology in this field often varies by discipline, we clarify below how VHAs [98, 176] relate to and diverge from other established agent concepts. VHAs represent a therapeutic specialization within the broader socially interactive agent research domain, incorporating elements from the following approaches:

- **Intelligent Virtual Agents (IVAs):** These serve as the foundational technology for VHAs. IVAs are defined as interactive digital characters capable of real-time perception, cognition, and action within dynamic social environments [105]. VHAs extend this foundation by specializing these autonomous capabilities specifically for the delivery of evidence-based clinical interventions.
- **Virtual Humans:** Often defined as artificial characters that "look and act like humans but inhabit a simulated environment" [183], this field frequently emphasizes high-fidelity biological realism [105]. By focusing on the "agent" rather than the "human" aspect, VHAs emphasize social intelligence and task-specific support without the necessity of photorealistic simulation.
- **Embodied Conversational Agents (ECAs):** Traditionally described as cartoonlike characters that demonstrate human-like properties in face-to-face conversation [26], ECAs provide the multimodal framework for coordinated verbal and non-verbal dialogue. VHAs advance this tradition by integrating these conversational capabilities into high-fidelity 3D environments, ensuring the real-time performance required for engaging health applications.
- **Social Robots:** While VHAs share the goal of engaging users on both cognitive and emotional levels to provide support [105], they are distinguished by their digital nature. Operating within 3D graphics software rather than physical hardware allows VHAs greater accessibility and scalability for widespread clinical deployment.

It should be noted that the term **Avatars** refers to virtual puppets controlled and tele-operated by their users; they do not require implementation of complex autonomous agency necessary for the implementation of the autonomous agents mentioned above. Unlike Virtual Health Agents, standard **Chatbots and Large Language Models (LLMs)** are typically disembodied, lacking the spatial presence and non-verbal signaling capabilities afforded by a 3D digital form. For the remainder of this paper, the terms "virtual health agent" and "virtual agent" are used interchangeably to refer to the 3D-embodied autonomous intelligent agents developed for this therapeutic context.

2.2 Related Research on Virtual Agents

The development of virtual agents as counselor extenders requires sophisticated coordination between verbal and nonverbal behaviors to establish therapeutic rapport and maintain user engagement. This section reviews existing approaches to multimodal behavior generation in virtual agents, with particular focus on active listening implementations and their application to therapeutic contexts.

2.2.1 Machine Learning Approaches to Multimodal Coordination. A much cited virtual agent data-driven active listening system is SimSensei [39]. It integrates multimodal perception with real-time behavioral generation, and uses a multimodal perception system to detect user non-verbal behaviors, including eye gaze patterns, head orientation, vocal tone variations, and conversational pauses [150]. These cues are used to drive the interactive behavior of the virtual agent (for turn-taking, listening feedback, rapport gestures) [104]. However, because matching the subtlety and timing of human-level nonverbal sensitivity is a very challenging problem, only 27.5% of participants in an evaluation study “agreed or strongly agreed” that SimSensei correctly sensed their nonverbal behavior. Furthermore, the system’s listening behaviors is largely reactive, rather than portraying active listening strategies coordinated with specific therapeutic communication goals.

Recent advances in deep learning have enabled more sophisticated approaches to speech-gesture coordination, which employ neural networks to achieve temporal synchrony between verbal and nonverbal modalities [92]. These data-driven approaches learn coordination patterns from large corpora of human interaction data, potentially capturing subtle relationships that rule-based systems might miss. Speech-driven gesture generation systems now use deep learning data-driven methods by incorporating representation learning in which nonverbal models use speech as input to produce gestures as output, as a sequence of 3D coordinates [200]. Similarly for facial animation, a study demonstrates that machine learning techniques can be used to drive 3D facial animation by audio input in real time with low latency [83].

While machine learning methods do well at capturing statistical patterns in multimodal data, they face significant challenges, however, in therapeutic contexts where behavior appropriateness and interpretability at all times are critical. Current machine learning systems lack the transparency needed for reliable therapeutic applications, making it difficult to ensure ethically sound, and clinically relevant behaviors. The data requirements for training therapeutic interaction models, furthermore, represents a significant bottleneck.

2.2.2 Rule-Based Multimodal Behavior Systems. Rule-based approaches, on the other hand, maintain explicit control over behavior generation, and enable designers to encode specific therapeutic principles. Early work in coordinated multimodal behavior generation established fundamental principles for virtual agent design: the Behavior Expression Animation Toolkit (BEAT) [27] provided the first systematic approach to coordinating speech and gesture through linguistic analysis, using rule-based mappings derived from observational studies of human conversational behavior. BEAT performed automatic linguistic and contextual analysis of input text to generate synchronized nonverbal behaviors, including facial expressions, head movements, and hand gestures.

The SAIBA (Situation, Agent, Intention, Behavior, Animation) framework [88] established a three-stage model for multimodal behavior generation: intent planning, behavior planning, and behavior realization. This architecture has the advantage of separating high-level communicative goals from low-level animation control, thereby enabling systematic rule development. Adaptive virtual agents have evolved beyond static response systems to implement variable behavioral patterns based on real-time user state assessment [12]. The Sensitive Artificial Listening Agent project, for example, computes listening behaviors that vary according to simulated agreement states, emotional traits, and interaction history [166].

These rule-based and data-driven foundational systems, however, were developed for general interaction and lack the specialized behavioral repertoires necessary for effective therapeutic communication. In addition, because systems using automatic recognition of user’s social cues (e.g. facial expressions) introduce privacy issues, it is not clear that most users are ready to accept and adopt them in a health-related context, as our study hints at.

2.2.3 Therapeutic Virtual Agents and Relationship Building. Foundational work on relational agents established critical principles for long-term human-computer relationships, demonstrating that virtual agents must maintain consistent personality traits, remember previous interactions, and employ appropriate social relationship models to sustain engagement over extended periods [16]. Their research identified key factors including social presence,

trust development, and gradual deepening of relational intimacy as essential components for effective virtual counseling relationships. However the study was focused on exercise adoption rather than therapeutic interventions, lacked evidence-based clinical frameworks like Brief Motivational Interviewing, ignored therapeutic environmental design considerations in which the agent is situated, and was evaluated with users in a non-clinical context, rather than addressing the complexities of real-world deployment to diverse clinical populations.

Recent therapeutic applications employ reciprocally adaptive behaviors that respond to user emotional states and modify interaction patterns accordingly [195]. Evaluation in cognitive behavioral therapy (CBT) contexts has shown that adaptive agents provided enhanced user experience and improved therapeutic effectiveness compared to non-adaptive agents. However, the study uses CBT as a proof-of-concept without implementing a complete evidence-based intervention such as Brief Motivational Interviewing, and does not provide comprehensive analysis of clinical populations' acceptability and diversity needs.

Recent reviews of the literature on virtual agents [102], and on health-related virtual agent in particular [180], also revealed that most research on virtual agents has been performed on the verbal behavior categories of speech or textual output, and on the nonverbal behavior (NVB) categories of the facial expressions, eye gaze, and nodding head movements sub-categories. We provide details later in this article on how we applied these design principles to manipulate facial expressions, eye gaze, and nodding head movements that have been shown to provide VHAs' abilities to simulate rapport [58], relational immediacy behaviors [14], empathic communication [98], and the ability to listen sensitively [167]. However, whereas these well-studied cues are essential to establish rapport and communicate empathy, counselors communicate warmth and caring using additional under-studied social cues, including the visual cues provided by their attire, and their environment design [125, 139, 172]. A study investigated the impact of a virtual counselor's environment design features and identified a room with a bench, a sink, and medical equipment as most appropriate and professional to deliver counseling [138]. Authors, however, selected these three objects based on objects typically present in Google images of examination rooms rather than of counselor's offices. As such, the work does not account for the "soft room" design features that counselors use to support the psychological and emotional aspects of counseling. Furthermore, the study recruited Amazon Mechanical Turk crowdsourced workers to evaluate the medical room design, and did not capture the perceptions of targeted health-afflicted populations in the real world, limiting the applicability of the results to real-world assessment of virtual agent counselors. Additional related research on VHA literature is provided in Appendix B Table 18, which we refer to in the following sections on agent social cues. Research methodologies used in these studies and guidelines from these are also shared to contextualize what has already been presented, and to illustrate additional opportunities for exploration.

2.2.4 Active Listening, Empathic Behaviors in Virtual Agents. Effective therapeutic virtual agents require sophisticated implementation of three fundamental behavioral categories: active listening behaviors that demonstrate attentiveness and responsiveness to user communication [127, 151, 192], empathic behaviors that recognize and appropriately respond to emotional states [46, 49, 56, 77], and immediacy behaviors that reduce psychological distance and enhance perceived closeness [91, 115, 124].

While these behavioral categories share overlapping characteristics, they serve distinct communicative functions essential for therapeutic rapport and sustained user engagement. Active listening, empathy, and immediacy all rely on behaviors that signal attention, warmth, and engagement, including eye contact, nodding, responsive facial expressions, and supportive vocal tone. Active listening centers on accurately tracking content and meaning, using behaviors such as paraphrasing, clarifying questions, and timed backchannel cues. Empathy focuses on recognizing and responding to another's emotional state, using behaviors such as validating expressions, compassionate tone, and affective mirroring (a.k.a. parallel empathy). Immediacy emphasizes reducing psychological distance through approach behaviors such as forward lean, open posture, close proximity, and expressive warmth.

Research on virtual agents demonstrates that listening behavior can be generated from features available in real-time, including understanding of spoken utterances, speech quality, posture shifts, and head movements [107]. Virtual agents implementing active listening behaviors have been perceived as more trustworthy, friendly, understanding, and socially attractive, with active listening and reassurance qualities identified as essential for authentic social behaviors that support engagement [79, 134, 191].

Empathic behaviors in virtual agents refer to the capacity to recognize, understand, and appropriately respond to user emotional states, encompassing cognitive empathy (understanding emotions) or affective empathy (sharing or mirroring another person's emotional state, often through automatic emotional resonance) [136]. Current state-of-the-art virtual agents simulate *affective empathy* using mirroring mechanisms that generate affective resonance with users' emotional states, using inductive affect-mirroring approaches (e.g., McQuiggan et al.'s parallel empathy) [114, 162]. *Cognitive empathy* is modeled through user-state inference, intent interpretation, belief-desire-intention reasoning and theory of mind [142] that enable agents to understand the meaning and context behind users' expressions, supported by theory-of-mind-oriented computational models [142], and on architectures for modeling emotions [18, 21, 108, 132, 153, 197]. Research reveals that agents exhibiting concurrent verbal and nonverbal empathic behaviors are rated highest on friendliness, warmth, and charisma, with inconsistency across empathic channels negatively affecting agent perception [15].

The On-Demand Virtual Counselor (ODVIC) system implemented active listening, empathy, and immediacy through multimodal perception of emotions and head movements, facial expression mimicry, reflective listening techniques, and continuous nonverbal engagement [98]. The empathic version achieved 31.36% higher intention to use than the same system delivered with a text-only interface, and 24.77% higher perceived empathy, with significantly higher ratings for trustworthiness and social presence than a non-empathic agent. Interestingly, a non-expressive agent was less accepted than a text-only interface, demonstrating that inappropriate multimodal behaviors can be counterproductive. However, the study was not conducted with clinical population, and did not investigate whether participants would prefer a system without automatic facial expression recognition.

2.2.5 Addressing Research Gaps. In summary, virtual agent counselor development faces critical research gaps despite advances in multimodal behavior generation. Current systems lack integrated frameworks that coordinate asking, explaining, and listening as unified communicative acts for active listening, instead treating these as separate behaviors. Research has insufficiently investigated nonverbal behavior timing and sequencing for therapeutic contexts, concentrating on general rather than clinical communication patterns. Additionally, there's an absence of design recommendations for a counselor's office environment as healing space, with existing research emphasizing medical settings that do not necessarily align with therapeutic office features. Finally, the field lacks taxonomic descriptions of virtual agent counselors' single and combined social cues, hindering systematic knowledge sharing across research labs and interdisciplinary fields. While machine learning approaches show promise for capturing complex multimodal patterns, they currently lack the interpretability and controllability required for robust therapeutic applications. Rule-based systems provide the transparency and explicit behavior control for safe clinical contexts, but existing rule-based frameworks have not been systematically applied to therapeutic speech act coordination.

Our work addresses these limitations by developing a virtual counseling system that combines evidence-based interventions, systematic active listening behaviors, and a therapeutic office environment design in a web-accessible platform. Our approach translates established therapeutic communication principles into specific nonverbal behaviors for each communicative acts: pairing initial greeting with theory-based proxemics cues, strategic questioning with supportive gestures, ongoing listening with embodied attention cues, and explaining with emphasis movements. In addition, we introduce the first therapeutic "soft room" environment specifically designed for virtual counseling, moving beyond traditional medical examination room metaphors to create a counselor's office space that supports emotional safety and therapeutic engagement. The system is evaluated with

the actual target clinical population (heavy-drinking adults) recruited through real-world online social networks, providing ecological validity that extends beyond controlled laboratory settings or crowdsourced participant studies. Our design recommendations are therefore based on evidence from counseling literature, and on findings derived from use of the system target populations that endorsed our design decisions.

2.3 Brief Motivational Interviewing Interventions

As explained further in Section 4, the virtual agent counselor we implemented delivers a digital evidence-based and patient-centered brief motivational interviewing intervention (BMI), targeted on alcohol use disorder (AUD). BMIs are based on Motivational Interviewing (MI) [121], an effective and scalable intervention for a variety of target behaviors involving changing lifestyle, e.g. alcohol use disorder (AUD), smoking, low vegetable intake, narcotics use, physical inactivity, non-adherence to treatment, among others [159]. It is a client-centered counseling style that is directive, and elicits behavior change by helping clients explore ambivalence and resolve it in order to develop *intrinsic* motivation to change. MI core principle is to amplify the discrepancy between a person's goals and their current behavior, with the idea that increased discrepancy heightens ambivalence, which can ultimately lead to change when resolved.

Adaptations of MI have bourgeoned to meet the need for motivational interventions that are brief [45], so that they can not only be used in counseling settings, but also in primary care settings. BMI interventions integrate MI's communication principles - expressing empathy, developing discrepancy, rolling with resistance, and enhancing self-efficacy - with core elements of effective brief interventions, including personalized feedback, change options tailored to level of readiness to change, and supportive guidance. BMIs demonstrate versatility across settings, functioning both as a standalone treatment (with even five minutes single-session interventions sometimes matching the efficacy of multiple sessions [8]), and in combination with other approaches including cognitive-behavioral therapy. Sustained benefits for long-term healthy lifestyles [44, 50], and effectiveness are supported by meta-analytic evidence [22, 66, 159].

Multiple barriers impede treatment access, including treatment need recognition [168], provider availability and training [190], geographic proximity, transportation access, cost, insurance coverage, scheduling flexibility, and stigma [186]. The widespread and escalating prevalence of these behavioral health issues worldwide [135] - exacerbated by the increasing number of mental health therapists experiencing burnout [23] - compounds the workforce shortage, creating a vicious cycle.

These barriers disproportionately affect individuals in rural areas and those experiencing poverty - populations where alcohol abuse prevalence is higher [84, 145]. Rural settings present additional challenges through reduced anonymity [145, 182], intensifying help-seeking stigma. Moreover, when treatment is accessed, it may not be evidence-based, creating additional quality care barriers. Indeed, the limited availability of locally trained personnel in BMI often restricts access to effective interventions for those in need [118]. Strategies to rapidly expand access to effective behavior change interventions across diverse populations are needed. Despite high alcohol use prevalence and the existence of effective interventions for AUD, for example, less than 10% of individuals with AUD in the USA receive any treatment [29].

To address this accessibility and stigma challenges, substantial evidence supports the overall efficacy of computer-delivered interventions that can extend clinician's practice as "clinician extenders", by providing a way to deliver high-quality, standardized evidence-based versions of screening, evaluation, and brief treatments [13, 24, 34, 74, 143, 173]. Despite their promise, however, the effectiveness of computer-delivered interventions hinges on maintaining user engagement and retention - ensuring that individuals remain motivated to interact with these systems and return for booster sessions, follow-ups, or ongoing lifestyle support as needed. Although computer-delivered BMIs have shown positive outcomes, their long-term impact is frequently undermined by high attrition rates, largely driven by users' low engagement during interactions [143, 189].

Our work leverages the ability of virtual agents to increase users engagement with computer-delivered interventions, with the goal to lower attrition rates, and increase long-term impact of computer-delivered evidence-based self-help behavior change interventions.

3 Theory-driven Virtual Agent Design Methodology

To design our virtual health agent system, we applied theory in three ways: (1) through a model on user's virtual agent acceptance (Section 5.1.1) that informed our design and future recommendations; (2) through an existing taxonomy of single social cues for virtual agents derived from social science literature [51] (Section 3.2); and (3) through principles of effective counseling gathered via literature search to inform metaphorical design of virtual health agent counselors and of their office (Section 3.3).

3.1 Virtual Agent Acceptance Model

Considering the multitude of social cues and features that a virtual agent can simulate, it is important to know which features are relevant or necessary to ensure user's acceptance of, satisfaction with, and ultimately adoption of a given technology. To get insights into the impact that various features of our virtual agent counselor system might have on user's acceptance, we used the model of Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC) by Van Vugt et al. [187], developed to inform creators of interactive characters. The model explains how user Satisfaction (itself an indicator of adoption) is shaped not only by the virtual agent's design features, but more importantly, by the user's *perception* of these features, emphasizing the need to investigate virtual agents from the user's perspective.

I-PEFiC combines engagement theory [87] with the Technology Acceptance Model [37] to predict how design variables influence users' acceptance of interactions with virtual agents. This model (shown in Figure 8 of Appendix C) distinguishes the Agent System - representing the intervention (e.g. quality of the BMI health content and implementation) - from the Agent Embodiment - with its social cues (e.g. facial similarity, realism). This separation facilitates focused discussions on the virtual agents' embodied features apart from the complete system implementation.

I-PEFiC distinguishes three user experience phases: *encode*, *compare* and *respond*. The *encode* phase involves user perception of: (1) the Agent System (a) Ethics perceived as trustworthiness (e.g. helpful vs. snooping), and (b) Affordances perceived as usefulness and ease of use; (2) the Agent Embodiment (a) Realism perceived as realistic vs. unrealistic appearance/behavior, and (b) Aesthetics perceived as attractiveness.

In the *comparison* phase, the user establishes personal Relevance (e.g. the agent is useful) and Valence (e.g. it's fun interacting with the agent) toward the virtual agent. The *response* phase concerns user Engagement, involving parallel tendencies of Involvement (rapport, empathy) and Distance (antipathy, irritation) toward the virtual agent.

The model predicts that both Use intentions and Engagement contribute to User Satisfaction: (1) for the Agent System, Affordances and Ethics are crucial for user Intentions and Engagement; (2) for the Agent Embodiment, Facial similarity, Aesthetics, and Realism contribute to Engagement, in decreasing order of importance. Embodiment Aesthetics and Facial similarity contribute to Use intentions in interaction with Agent System Affordances, while Realism does not contribute to Use intentions.

In this work, we discuss how the I-PEFiC model guided us in our choice of social cues and their design.

3.2 Social Cue Definition and Taxonomy

3.2.1 Definition of Social Cues. To describe in details the variety of social cues that our agent displays and the user's socio-emotional responses that they trigger, we use definitions of cue, social signal, socio-emotional response, and social cue from terminology discussed in the Computers As Social Actors (CASA) theory [148],

which posits that humans automatically apply the same social heuristics used in interpersonal interactions to computers, as these evoke similar social attributes to those of humans. We provide two examples in Figure 1: a *cue* is any design feature of a VHA noticed by the user that provides information (e.g. gender of the voice; a "soft room" design with pleasing decor and plants); a *social signal* is the interpretation (conscious or unconscious) of cues by the user in the form of attribution of mental state or attitudes about the VHA (e.g. attribution of a biological gender towards the VH; attribution of a safe therapeutic room toward the VHA office); a *socio-emotional response* is an emotional, cognitive, or behavioral response of the user toward a VHA that is considered socially appropriate when directed toward human beings (application of gender homophily toward the VHA; feeling comfortable in the office); and a *social cue* is a cue that triggers a *socio-emotional response* of the user toward the VHA, i.e. a room design features can be considered as social cues if they trigger socio-emotional responses, such as feeling comfortable in the room.

3.2.2 Taxonomy of Social Cues. Given the very high number of the variety of possible combinations of individual social cues that contribute to human-human social interactions within a physical environment – and that can therefore potentially be simulated by a virtual agent – using these definitions of social cues in Section 3.2.1, researchers have proposed an empirically validated taxonomy of social cues derived from a systematic review and analysis of literature on interpersonal communication to document agent designs. Feine et al. [51] classify single social cues at the lowest level of complexity into taxonomy categories, which provides a good starting point to accumulate knowledge about agent social cue designs from different research laboratories, disciplines and domains, and different contexts and cultures.

The taxonomy identifies 48 social cues, and these are organized in four main categories of social cues at the first hierarchical level - visual, verbal, auditory, and invisible - and ten subcategories at the second hierarchical level, where each subcategory includes a set of single social cues. In Section 5 we discuss the individual social cues we implemented in our agent, with our design rationale.

It should be noted that this low level description of single social cues does not ignore the fact that single cues do not occur in isolation during social interaction (e.g. speech needs to be congruent with facial expressions to elicit trust). In Section 5.1.4, we describe how we combined a collection of these single verbal, non-verbal, and proxemics social cues to model active listening communication, specifically designed for the context of delivering counseling for behavior change interventions. In Section 10, we summarize how we extended Feine et al. [51] original taxonomy (as suggested by its authors), in an effort to assist further systematic accumulation of detailed and systematic knowledge on the design of virtual agents' social cues.

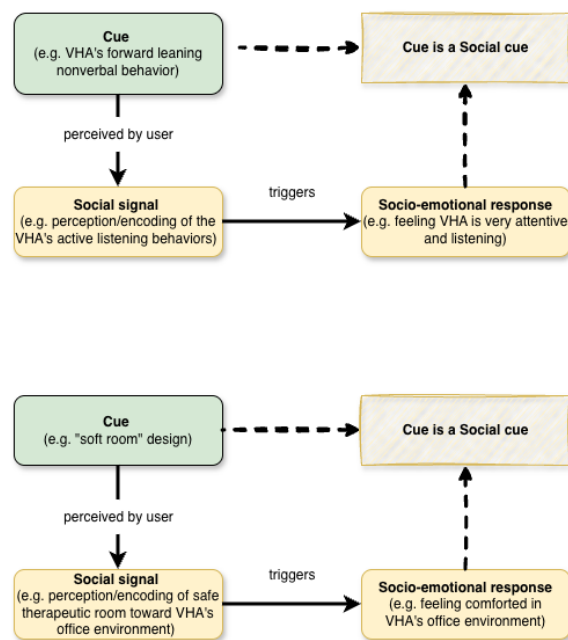


Fig. 1. The emergence of a social reaction towards a cue of a virtual agent defines a social cue (example adapted from [51, 126]): when a cue of a virtual health agent (VHA) counselor is interpreted as a social signal and triggers a socio-emotional response, the cue is considered a *social cue*. The top example shows the process for a VHA's forward leaning nonverbal cue; the bottom example shows the process for a VHA's "soft room" design cue.

3.3 Human Counselor Metaphor

Applied to the current context of health counseling, the Computers As Social Actors theory [148] predicts that if virtual agents are designed to provide social cues similar to those of a successful counselor, people will have similar socio-emotional responses to the ones they would have to a human counselor. Toward that end, we used *metaphorical design* which, in the case of agent design, provides information about morphology (e.g. human form), voice, verbal and nonverbal behavior, and other physical features related to the given context (e.g. clothing, environmental background) that support the desired metaphor [111].

We used the **metaphor of a successful human counselor practicing in an office with adults living in a Western culture** to inspire our design. Our target end-users are English-speaking adults living in the USA, who engage in heavy drinking (i.e., consuming 5 or more drinks in one sitting at least once in the past year), and are interested in getting support but not currently receiving services. We conducted reviews of the literature based on this metaphor, and a similar methodology could be applied to meet the needs of other target population by surveying literature on how counseling is conducted in other countries to derive relevant social cues and inform design.

Metaphorical design of a successful therapists is challenging because no precise set of rules exist about what makes a therapist better than another [28], and individuals can have their own preferences about what is desirable in a session with a human counselor [78, 144]. While therapeutic working alliance has emerged as an important variable for successful therapy or counseling - namely, agreement on goals, assignment of task, and development of bonds [17] - knowledge of the specific therapists' contributions to the development of the alliance are needed for the design of virtual agent counselors. We used research on counseling to identify these, which rather than pointing to personality traits found in the well-known Big5 five-factor model of personality [33], pointed us instead to therapist's personal characteristics.

The following **personal characteristics of successful counselors** have been identified in a review of therapist characteristics across a variety of therapeutic orientation: *warm, friendly, open, interested, experienced, trustworthy, respectful, confident, flexible, alert, honest, and attentive* [2]. These are the characteristics that we aimed would emerge from the various features we included in our VHA design, so that they are perceived by our users, and trigger the same socio-emotional responses they would have in a therapy session with a counselor.

The challenge, however, is that having these personal characteristics relies upon the expert use of a large number of combinations of verbal and nonverbal social cues for effective active listening skills used to establish and maintain a positive therapeutic alliance. As we discuss in Section 5, these verbal and nonverbal active listening communication abilities are furthermore combined with the counselor's appearance, and implicit social cues embedded in the counselor's office as a healing place. Across counseling and psychotherapeutic mental health interventions, the ability for the counselor to provide a comforting environment where the patient can engage and feel safe is essential [1]. Human counselors develop these characteristics through a complex blend of science and art [42, 70], creating substantial challenges for virtual agent counselor implementation.

Our approach was aimed at eliciting comfort and familiarity by invoking the user's mental model of the VHA's abilities based on users' expectations of the characteristics of a successful therapist mentioned above, by implementing an evidence-based intervention that the VHA delivers, designing individual cues (e.g. clothing style, room decor, body language, verbal tone), and combining them to implement well-received VHA's active listening behaviors, and a comforting VHA's office environment.

Our design process is therefore analytical, and theory-driven: the agent's metaphor is based on social psychology theoretical models of human counseling communication, evidence-based health interventions, and principles of human-centered design. In the following Sections 4 and 5, we discuss how we used both established guidelines and novel agent features for our VHA design based on the human counselor metaphor.

4 Design of Health Intervention Virtual Agent System

The virtual agent health intervention system, named eEVA-4AUD, is web-based and requires only a browser to launch from its URL, ensuring accessibility. The landing page is the Counselor Selection Menu shown in Figure 2. The agent is embedded in a software system that manages the agent's behaviors and database.

The virtual health agent acts as the main multimodal user interface, conversing in real time with the user within its virtual office environment: it speaks and listens to spoken answers, displays congruent facial expressions and body gestures, and stores user answers in a database. It is complemented by traditional UI widgets (e.g. menus, buttons, checkboxes, text boxes, bar graphs) to provide system feedback and facilitate visualizations when needed, applying well-known usability principles.

In the following, we discuss the complete evidence-based brief motivational interviewing content implemented in the virtual agent system, and provide an overview of the main system components with additions made to increase engagement, including how we integrated traditional UI elements to ensure robustness, usability, and positive user experience.

The detailed design of the virtual agent single and multimodal social cues used to simulate counselor characteristics is discussed in Section 5.

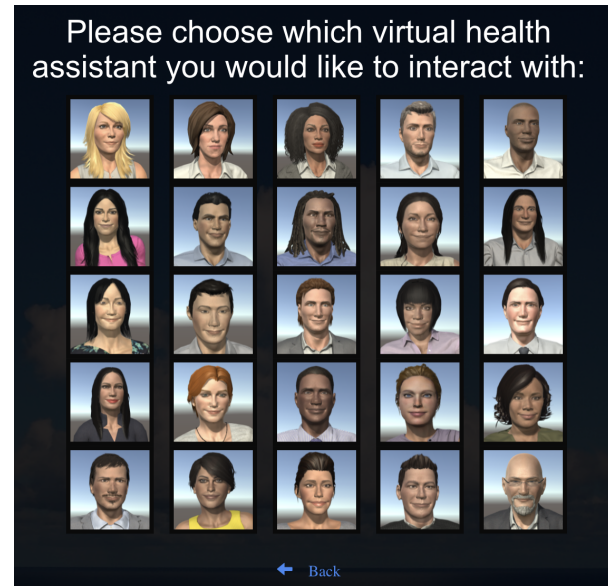


Fig. 2. Counselor Selection Menu: If the user hovers over on one virtual agent counselor (VHA) image, a closeup of the VHA is shown with its name; when the user clicks on a VHA, the VHA starts interacting in its office (e.g. Fig. 4).

4.1 Content of the Brief Motivational Interviewing Intervention

The health intervention that the agent delivers is based on the structure and verbal content of the Drinker's Check-Up (DCU), an evidence-based approach incorporating Motivational Interviewing strategies (see Section 2), that specifically targets excessive drinking behaviors. We chose the DCU because it is a complete BMI, and we could therefore engage our target users, over multiple sessions if users wanted, with a helpful health intervention. The DCU is evidence-based, hence we could safely offer it to a sample of our heavy drinkers target population because it has been amply validated through hundreds of clinical trials [25]. As discussed in Sections 6, 7 and 8, conducting the study with a sample of our target population enabled us to gather meaningful impressions from the types of users we built the system for.

The DCU contains three sequential sections - Assessment, Feedback, and Change Plan - that the user can complete in one sitting or more:

- (1) *"Look at your Drinking"*, where users first share comprehensive drinking information including behavioral patterns, and family history;
- (2) *"Get Feedback about your Drinking"*, where users then receive personalized and empathic feedback highlighting risk factors and consequences based on results from standardized assessment measures and the user's age and gender group (examples of the UI for the Feedback section are shown in Figure 3);

(3) "Design an Action Plan to Change", where users work through ambivalence about change via balance exercises, develop action plans, and establish goals toward change based on their current readiness to change (the UI that assesses readiness is shown in Figure 4).

At the end it offers an optional printable summary of the tailored feedback information. Importantly, the DCU maintains the user's autonomy by not telling them what to do: the user alone decides whether they want to change their alcohol use. Research demonstrates DCU's effectiveness, showing 50% reductions in drinking quantity and frequency over 12 months compared to controls, establishing it as one of two web-based digital health interventions with the strongest RCT evidence [24, 25].



Fig. 3. Examples of the user interface of the virtual agent eEVA-4AUD system for the *Get Feedback about your Drinking* Module. The virtual agent interaction is complemented on the right side of the user interface (UI) by traditional UI elements, e.g. bar graphs, tables, to add visualizations to the spoken verbal feedback, minimizing user's memory load with the recognition rather than recall usability principle.

4.2 Verbal Additions to the Original BMI Intervention and Traditional User Interface Elements

The virtual health agent system we implemented - eEVA-4AUD - uses the VHA to deliver the DCU, with additions to increase the realism of the counseling session metaphor, user's engagement and satisfaction. The agent talks and listens to the user's spoken utterances, or takes keyboard entries if the user prefers to answer without speaking out loud. Natural language understanding is limited, and not powered by LLMs that are not yet safe for deployment of health interventions to vulnerable population, due to potential harmful hallucinations. To render the conversation more natural than fully scripted interactions, we used keyword pattern matching to enable the agent to understand partial answers, synonymous terms, or accept slightly different nuanced user's affirmative responses, e.g. "yes", "sure", "ok".

We added the following content to the DCU, by combining guidelines we derived from our survey of social science literature, design principles for virtual agent design discussed in Section 2 [14, 35, 89, 98, 102, 165, 180], well known design guidelines for user experience [64], usability heuristics [128, 129], and design guidelines specific to human-AI interaction [6]:

- a **greeting scenario** by the VHA to virtually approach the user, similarly to what counselors would do to welcome patients into their office in a physical setting, to match relevant social norms [6, 85], shown in Table 1 and Figure 6;
- a **self-introduction** by the VHA explaining that it is a research prototype with some autonomous (albeit limited) social intelligence, with expertise on delivering an intervention for AUD vetted by the *US. National Institute of Health*, to clarify the user's expectations about what the VHA can do, and how well it can do it [6];
- **traditional graphical UI elements** to ensure high usability [129] by (a) providing visibility of system status (e.g. check boxes that get checked with user's answers to provide immediate feedback that agent understood answer), (b) offering flexibility and user control (e.g. check boxes that users can check manually if they do not want to speak their answers, or if they have difficulties with the speech recognition feature), (c) facilitating error prevention with a 'go back' verbal feature, and (d) providing easy to understand efficiency of use with visualizations of data (e.g. bar graphs, tables); UI examples are shown in Figure 3.



Fig. 4. Example of eEVA-4AUD Graphical User Interface: The agent chosen from the character menu (cf. Figure 2) loads in the 3D office environment. The agent goes through the content of the eEVA-4AUD three main sections. This page shows the agent asking about the user's readiness to change drinking patterns at the start of "Design an Action Plan to Change" section. Users answer by speaking in a microphone ('mic icon' is turned on), or by clicking checkboxes using a mouse if they don't want to speak up or have issues with their microphone; the appropriate checkboxes get marked to provide system status and the user can ask to go back to correct the answer for error prevention; in other sections the user can enter text with a keyboard. Word pattern matching is used for the agent to understand partial answers. From any screen the user can view *My Progress*; access *FAQs* about the system, or view/update his/her *User Profile*.

- **an optional tour** of the system offered by the VHA on the first encounter, where the VHA explains the basics of how to interact with it (e.g. turn taking indicated with on/off microphone icon, close captioning) to provide flexibility and efficiency of use [129];
- **addressing the user by the name** they chose to provide, to personalize further the interaction with the VHA [40], and to match relevant social norms [6];
- **an explanation upon the resumption of an interrupted session** that the VHA remembers the user's previous answers, and where they left off in the previous session, to reassure users no information was lost between sessions;
- **verbal acknowledgements and affirmative backchannel responses** congruent with nonverbal cues (see Section 5.1) to portray active listening behaviors, build rapport, validate the user, and manage turn taking with speaker transitions signaled with on/off microphone icon;
- **a farewell scenario** after the user has completed all the components of the eEVA-4AUD intervention, with best wishes moving forward to match relevant social norms [6].

The intervention is therefore a lengthy evidence-based intervention with many steps that users can complete at their leisure. A sample excerpt of the agent dialog with a user is provided in Table 1, showing the stages of the intervention and the applied design principles or the relevant BMI feature. Additional technical details on the implementation of the system are available [19]. Interactional freedom is limited as in the original DCU to ensure consistent content delivery. Given that users may not interact with every portions of the intervention (as it is tailored to each user), and that they can stop, repeat or return to sections of the intervention at their leisure, the interactions between the human and the virtual agent are variable in duration. If the user does not complete a session within one sitting (as some users did), the system saves the answers; upon returning to the system, the VHA lets the user know it remembers where they left off their interaction last time, and resumes the interaction from that point on. In this study, we asked participants to fully complete the three sections of the intervention within a month.

5 Metaphorical and Taxonomic Design Features for virtual agents as Counselors

To design our virtual agents as counselor extenders within the healthcare ecosystem, we used both existing and new design guidelines: we used established design guidelines for the most studied social cues, and new guidelines for under-studied social cues such as VHA's attire and office environment that we derived from counseling literature. We provide a very fine-grained comprehensive description of these that is not often provided (see Section 2), using Feine's taxonomy of social cues [51] in an effort to systematize and document our design rationales and their implementation with high level of details to enable research replication and improvements.

To discover how to design visual, verbal, auditory, and invisible social cue features that have been less studied but necessary for our effective human counselor metaphor, we conducted a survey of the social science literature on counselors', psychotherapists', and clinical psychologists' social cues, to extract design features that tend to trigger desired socio-emotional responses in real counseling sessions.

As discussed in Section 3, we aimed to simulate **personal characteristics of effective counselors** successful across different types of therapies: namely, *warm, friendly, open, interested, experienced, trustworthy, respectful, confident, flexible, alert, honest, and attentive* [2]. Our design choices for individual social cues were made with the goal that users would perceive them and interpret them as personal characteristics of the virtual agent counselors, and in turn trigger similar socio-emotional responses to those they would have in a therapy session with a human counselor.

In the following subsections, we detail our choice for every single social cue, and combinations of, that we included in our VHA design with a rationale for our choices, using the terminology, definitions, and examples provided in Feine's et al. taxonomy of social cues [51]:

Speaker	Dialog Excerpt	Health Intervention Section	Design Principle or BMI feature
Agent:	- Hello, my name is Luke, a virtual agent designed to talk to you about your health. Welcome Joe!	Greeting	- Make clear what the system can do & Match relevant social norms & User name to personalize intervention
	- I do not have full human social intelligence, but I can be useful as a clinician extender by delivering a brief health intervention vetted by the National Institute of Health about alcohol use disorder.	Self-introduction	- Make clear how well the system can do what it can
	- Is it OK if I ask you a few questions about your drinking, Joe?	Asking for permission to move to next section	- Respect autonomy, build therapeutic alliance
User:	- Yes, go ahead.		
Agent:	- It's a good decision Joe to take a look at your drinking. Did you ever in your life felt bad about yourself because of your drinking?	Look at your Drinking	- Praise, build motivation
User:	Actually, yes I have.		
Agent:	- I see.		- Acknowledge, build therapeutic alliance with active listening (verbal & nonverbal)
	- And in the last 3 months, how often has it happened: never, once or a few times, once or twice a week, or daily or almost daily?		
User:	- Well, once or twice.		
Agent:	-Okay, thanks.		- Acknowledge, thank, build therapeutic alliance with active listening (verbal & nonverbal)
...	...		
Multiple turns	... [Agent continues to ask questions for user to answer]. . .		
...	...		
Agent:	Very good, you have now completed the Look at your Drinking session.		- Praise, acknowledge progress
	Is it OK if I give you feedback about your drinking, Joe?	Asking for permission to move to next section	- Respect autonomy, build therapeutic alliance
User:	Yes please.		
Agent:	The number of standard drinks you consume in a typical week is 58, which means that you drink more than 97% of American males. Is this number lower than you expected, higher, or about what you expected?	Get Feedback about your Drinking	- Provide tailored normative feedback.
User:	Wow, it's much higher!		
Agent:	Oh, I see.		- Acknowledge, build therapeutic alliance with active listening (verbal & nonverbal)
	It can be surprising and (for some) discouraging to see that they fall higher on the alcohol use scale than they expected. Some people think there might be some mistake in the way results were calculated. Feel free to review your answers. Also, it is not uncommon for people to socialize with others who drink the same as – or more – than they do. Think about which (if any) of these reactions applies to you...		- Empathic feedback acknowledges user's potential emotions
Multiple turns	... [Agent continues to ask questions for user to answer]. . .	Asking for permission to move to next section	
Agent:	Before we consider what, if anything, you might choose to do about your drinking, Joe, it would be helpful to know how ready you are to change. Please choose from the following options (see options shown in Fig. 4).	Design an Action Plan to Change	- Determine readiness to change
User:	I have no interest at all in changing my drinking and I'm not interested in considering any changes in the future even if my situation changes.		
Agent:	It is certainly your decision to change or not to change your drinking. Before you leave, would you like to see how others who are thinking about changing go from here? [interaction continues]		- Build intrinsic motivation vs. external compliance, reduce resistance, respect autonomy

Table 1. Excerpts from agent’s dialog with a participant (column 2) for each Brief Motivational Interviewing (BMI) intervention section (column 3), along with associated virtual agent design principles, or BMI features (column 4).

- (1) in Section 5.1 we discuss **Visual** social cues we implemented, which refer to cues that can be visually perceptible (except words themselves);
 - (a) in Section 5.1.1 we discuss visual **Appearance** single social cues we implemented, which refer to an agent's graphical representation;
 - (b) in Section 5.1.2 we discuss visual **Kinesics** single social cues we implemented, which refer to all body movements of the agent;
 - (c) in Section 5.1.3 we discuss the visual **Proxemics** single social cues we implemented, which refer to the role of space, distance, and territory in communication;
 - (d) in Section 5.1.4 we discuss how we **combined Kinesics and Proxemics** single social cues to model **Multimodal active listening** behaviors;
 - (e) in Section 5.1.5 we discuss the visual **Computer-Mediated Cues (CMC)** single social cues we implemented, which refer to visual elements that can augment or modify the meaning of a text-based message;
 - (2) in Section 5.2 we discuss the **Auditory** social cues we implemented, which refer to cues that can be heard;
 - (3) in Section 5.3 we discuss the **Invisible** single social cues we implemented, which refer to cues that cannot be seen or heard
 - (4) **Verbal** social cues, which refer to cues expressed with written or spoken words;
- In Section 10, we discuss how we extended this taxonomy for future VHA designs.

5.1 Design Features for **Visual** Social Cues

According to Feine et al.'s [51] taxonomy that we use to describe our virtual agent [51], Visual cues refer to cues that can be seen (except written words themselves). Individual cues within this category are categorized into the following sub-categories of visual cues: **Agent Appearance, Kinesics, Proxemics, and Computer-Mediated Communication (CMC)**. Each of these sub-categories includes multiple single cues, which we expanded upon (see Section 10).

In the following subsections, we describe each of the Visual social cues that we implemented in our agent, using Feine et al.' [51] taxonomy terminology and definitions, and literature-informed rationale for our design choices.

5.1.1 Design Features for Visual Appearance Cues. The visual Appearance social cues are broken down into 10 single cues: 2D/3D, (degree of) human-likeness, (photo) realism, facial feature, attractiveness, age, gender, and color/race demographics, clothing, and name tags. In what follows, we specify how we implemented each of these Appearance cues, based on literature-informed decisions.

2D/3D Social Cue. The 2D/3D social cue refers to the *2D/3D agent appearance as the agent being displayed in either 2D or 3D*, e.g. 2-dimensional agent, or 3-dimensional agent.

Virtual agents' appearance can be achieved via multiple modalities, ranging from *graphical 2-dimensional (2D)* line-drawn cartoon- or comic-like characters, to virtual 3D graphics animated characters rendered in realtime during interactions. As pointed out in a recent review of the literature [180], most current agents for the health domain have been implemented as stylized cartoon- or comic-like two-dimensional (2D) characters (referred to as embodied conversational agents rather than virtual agents because of their 2D rather than 3D embodiment, as discussed in Section 2), with a few exceptions that have leveraged the 3D nature of virtual agents [39, 98, 147].

From a human-centered design perspective, however, compared to 2D graphics, objects in 3D graphics have been found to significantly increase three important aspects for their users [199]: *usability*, in terms of user satisfaction and decision confidence; *positive overall user experience (UX)*, in terms of perceived ease of use and

usefulness; and *sense of presence*, in terms of the subjective sense of being involved with the digital environment [154].

Since 3D graphics increase positive UX and usability (which include easy of use and usefulness) compared to 2D, and since the I-PEFiC model predicts that perceived ease of use and usefulness are strong predictors of use intentions, we designed our characters using 3D graphics, shown in Figures 2, 4, 5, 6. We also designed the environment in which they evolve using 3D graphics to increase the user’s sense of presence with agent in its office environment (see Section 5.1.3).

Humanlikeness Social Cue. The degree of human-likeness social cue is defined as *the appearance of the agent that ranges from a natural human form to the shape of a cartoon or artificial object*, e.g., the virtual agent has a human-like, object-like, animal-like, or robotic-like appearance. Agents’ appearance that represents best the live humans they aim to simulate have been found of high value in research on the impact of virtual agents’ role [10]. Given that we used the metaphor of human counselor for adults, we designed our VHAs to be humanlike.

Realism Social Cue. Realism (referred to as Photo-realism in the original taxonomy) is defined as *the appearance of the agent that ranges from an extremely photo-realistic appearance to a comic appearance*, e.g. real photo of a human being, a comic figure.

Since the I-PEFiC model predicts that the level realism of the agent’s embodiment contributes to users’ engagement with the virtual agent (see Section 5.1.1), and since one of our main goals is to increase engagement with digital health interventions, we took careful consideration of the realism of our characters. As we discuss below, realism levels vary from low (cartoon-like) to photoreal, so we renamed the ‘photo-realism’ social cue to ‘realism’ to facilitate a discussion on the possible levels of realism. Research on the influence of characters’ realism on users’ experience in the education domain found that users learned better with realistic pedagogical agents versus cartoon-like ones [10].

However, the *uncanny valley hypothesis* [123] postulates that users’ response to a humanlike virtual agent abruptly shifts from *empathy* to *revulsion*, or *eeriness*, if it tries to portray a lifelike appearance but fails. Researchers,



Fig. 5. Examples of virtual agent counselor’s appearance cues, and of three facial expressions, left to right: neutral, happy/smile, concerned. The menu of counselors shown in Figures 2 uses the (middle) smile expressions to first show the agents, along with their name.

therefore, search for realism levels that avoid the uncanny valley. Rapid progress in photorealistic 3D graphics rendering has made it possible to generate characters that, on a 2D digital display, cannot be distinguished from a human being [75, 110]. One study revealed that participants had *high affinity* with photorealistic characters, but only *if* they thought these were controlled by a human; if controlled by an autonomous artificial agent, participants found that a perfect human form would “*creep you out*,” “*be more spooky*,” and *confuse* them because they would not be able to distinguish whether the face is that of a human, or of an artificial agent [169].

Our goal was to provide users with 3D characters rendered realistically enough to trigger users’ mental model of a human counselor, while avoiding confusing users about whether they were interacting with a real human or a digital one. We therefore designed realistic characters that are purposely not photorealistic, shown in Figures 2, 5, 6, and 4. To help accumulation of systematic knowledge on virtual agent design, we also document our rendering style as closest to the *HumanBasic* realism level using McDonnell’s 10-item realism scale [110]: the style is skin-textured and uses card-based hair, but does not use subsurface scattering, nor advanced effects for rendering hair and eyes. In Section 10, we provide the 10-item realism scale that can be implemented and rendered, scale that can be expanded upon based on computer graphics progress.

Facial feature Social Cue. The facial feature cue is *the static and non-changing characteristics of the agent’s face*, e.g. static smile of agent.

In our virtual counseling application, static images are used to introduce the entire collection of 25 VHAs to the user, in the first screen after login in, where they are asked to select the VHA they want to interact with (see Figure 2, and the middle image in Figure 5). To give a positive first impression, they are shown with a static positive facial expression - a slight smile - rather than a neutral expression.

Camera Angle / Viewing Angle. In order to account for the viewing angle of the facial feature, we added another social cue to Appearance in the taxonomy, which we define as *the horizontal and vertical angle from which the virtual counselor’s face is presented in the initial view*. In our study, the static images used to introduce the 25 VHAs to the user portray a full frontal (0°) direct face-on view.

Attractiveness Social Cue. The attractiveness cue is defined as *the visual characteristics of the agent perceived as being pleasing or beautiful*, e.g. symmetrical face of the agent.

Although rarely studied, therapists’ attractiveness can have an impact on aspects of the therapy, e.g., clients tend to *disclose* more to attractive therapists [63]. Findings suggest that adults with strong rural identities demonstrated more positive impressions of virtual agent attractiveness. For these patients, optimizing virtual agent clinician appearance is critical when tailoring cancer prevention intervention [32]. Furthermore, the I-PEFiC model predicts that attractiveness, perceived as the aesthetic quality of the virtual agent, in turns significantly impact user engagement, and satisfaction.

We therefore aimed for our VHAs to be relatively attractive. Face attractiveness perception has been found to be influenced by facial symmetry, and averageness for both sexes of face characteristics [106]. All characters’ faces were designed with symmetry, and averageness for both sexes, and shown in Figures 2, and 5. While our current designs are based on average perceptions of attractiveness in contemporary Western societies, it is important to note that attractiveness is subjective, era- and culture- dependent. Furthermore, although attractiveness also has dynamic properties, we did not take these into account in this study, apart from designing dynamic facial expressions that are also symmetric.

Demographics Social Cues. We discuss together the several appearance social cues related perceived to demographics identified and defined in Feine et al.’s taxonomy [51]. The *age* social cue is described as *the biological age of the appearance of the agent, such as old agent visualization (e.g., grey hair), young agent visualization (e.g., shiny skin, full hair)*; the *gender* social cue, is described as *whether the agent is perceived as belonging to either one*

of the two sexes (male, female) or it is ambiguous; and the *color* social cue is described as *the overall color of the appearance of the agent*, e.g. white, black, or pink.

Gender, age, and race can be communicated by visual appearance attributes (as well as by its voice, discussed later). Whereas race has been considered a social construct based on stereotypes of physical attributes and on labels which are often treated as distinct, race variability and overlap do exist in a population [80], as well as for gender identity, ethnicity, and age [90, 101].

Given that the homophily principle posits that similarity breeds connection and influences *closeness* in human relationships [113], it is an important concept in our healthcare context. Research in social psychology indicated that humans are more influenced by someone belonging to their demographic group [11]. In the counseling and behavior change healthcare context, teenagers envisioning a VHAs to assist their weight management indicated that they wanted virtual agents that they could identify with physically, emotionally, and ethnically or culturally [94]. Robertson et al. [152]'s study on an eCoach for colon cancer, also found that the VHAs' demographic appearance (e.g. gender, race, age) should align with the target user population's demographics. A recent survey of virtual agents found, however, that the virtual agents used in the literature lacked in diversity in terms of characteristics associated with gender, age, and race (or skin color), with 67% of virtual agents (out of 48 surveyed) being white-skinned [102].

Furthermore, in addition to the need for diverse healthcare personnel pointed to by the social science literature mentioned, the theoretical I-PEFiC model predicts that (facial) similarity influences use intentions. Given that our virtual agents are intended for adults living in the US, and that the diversity index in the US was 61.1% in 2020 per the US Census Bureau [185] - i.e. there is a 61.1% chance that two people chosen at random will be from different racial and ethnic groups - we aimed at providing users with a choice of diverse VHAs, and to determine whether they appreciated being given that choice.

While we did not seek to make absolute claims about how users would interpret the VHAs' demographic features by labeling our agents as belonging to a specific racial or ethnic group, we did strive to create the visually unique set of 25 unique VHAs shown in Figure 2, and Figure 5 closeups. The VHAs are morphologically diverse in terms of varying characteristics that are typically used to differentiate race, gender, and age. We iterated with our graphics team by manipulating jawline, eyebrows, nose, lips, complexion, cheeks, eyes, hair style, and age features. We tried to create unique characters by simulating a mix of morphological features that users might perceive as potentially having their own demographic identity. While we did not try to portray all categories used by the current US Census bureau (namely, American Indian/Alaska Native, Asian, Black or African American, Native Hawaiian/Pacific Islander, and Caucasian/White) due to lack of resources, we did try to portray American Indian, Asian, Black or African American, and Caucasian/White. We also tried to portray female and male sexes, and a range of ages ranging from Adult (26-39), Middle-Aged (40-64) to Older Adult (> 64 year-old) (see Figure 2).

Clothing Social Cue. Clothing is described as *the items that the agent wears, such as clothes and trousers, to cover, protect, or decorate its body*.

Clothing can heavily influence first impressions by communicating information and messages conveyed by attire. Impression-formation theory postulates that clothing has a clear influence on the traits attributed to the people one meets. A study illustrated that, based on the counselor's clothing, clients formed initial impressions of counselors in terms of *caring*, *authenticity*, *competence*, and *trustworthiness*, and tended to *agree* more with the information provided based on the counselor's clothing [68, 161]. Importantly, first impressions of a counselor's attire can affect the development of dyadic rapport [53], and the impressions during initial encounter often determines whether a client chooses to return and decides to remain in treatment [36]. In another study about counselors' clothing, both male and female clients preferred a moderate style of attire, neither too formal nor too casual, i.e. business casual. When asked to choose a counselor that the client would most likely want to see, 81% of the clients chose the ones wearing a moderate forms of attire: they gravitated towards male counselors

wearing a shirt, tie, and slacks, or shirt and slacks; and female counselors wearing a dress, or blouse and slacks [68].

Within our theoretical framework, clothing relates to the embodiment aesthetic and realism, which have both been shown to influence engagement (see Section 5.1.1) and is important for our goal of increasing user engagement with digital health interventions. As suggested in the literature for caring, authentic, competent, and trustworthy human counselors, we therefore designed VHAs' **attire as neither too formal nor too casual attire, i.e., business casual**: female-looking VHAs therefore either wear a dress, a blouse (with or without a casual jacket) and slacks, or a skirt with a sleek sweater; and male-looking VHAs wear slacks, with either a shirt, with or without a tie, and with or without an informal jacket (see examples in Figures 2, 4, and 6). We note that clothing (and hair styles) are subjectively attractive, era- and culture-dependent, and that we designed for contemporary Western[-ized] populations.

Name tag Social Cue. *The name tag refers to a tag or badge with the name of the agent.*

Research on how design features for web-based health promotion programs that promote immediacy suggest using the provider's name, when interacting with them to personalize communication (and the same is true about using the user's name) [86]. We therefore gave a name for each agent. In our design, users could view the VHA names when they hovered over the character's facial image in the character choice menu (shown in Fig. 2).

5.1.2 Design Features for Visual Kinesics Social Cues. According to Feine et al.' taxonomy we use to describe our virtual agent [51], Kinesics social cues belong to the main Visual main category of social cues, and refer to all body movements of the agent.

The visual Kinesics cues are broken down into 5 single social cues: facial expression, head movement, eye movement, arm and hand gesture, and posture shift. In what follows, we specify how we implemented each of these social kinesics cues, based on literature-informed decisions, and on a observational study of a video corpus of counselor-patient MI sessions.

With the goal to emulate a counselor delivering a BMI, we first videotaped sessions between a live counselor and a client. Then, a clinical expert reviewed the videotapes to code verbal reflections and NVBs. From these observations, the expert generated a set of rules for basal behaviors of the VHA, associating speech acts to different combinations of facial expressions, head nods, eye movement, posture shifts, arm and hand gestures. Five hand gestures congruent with facial expressions and speech acts, including head nods, forward leans were derived from the clinical expert's rules.

In Section 5.1.4, we discuss how we combined these kinesics single social cues with proxemics single social cues to create a multimodal model of communicate acts, specially appropriate for counseling.

Facial expression Social Cue. Facial Expression is defined as *the expression of gesture by executing one or more motions with its facial muscles* [51].

The face represents one of the most important modalities in human social communication [5, 99]. Therapists' high facial expressiveness, involving a composite of nods, frowns, and smiles, can express encouragement, and has been associated with liking, better patient health outcomes, and less patient confusion [4, 70, 109, 151]. Therapists' micro facial expressions in the first session have an impact on the relationship with patients [117].

Because the Facial Action Coding System (FACS) [48] (discussed below) has been used to assess therapists' abilities to recognize or display facial expressions at the level of FACS action units [70, 117], it is important for VHAs to be able to display FACS-based facial expressions:

- *Facial Action Coding System (FACS).* FACS classifies all visible facial and head movements observed in the human face and head by defining the smallest possible group of muscles that can cause facial or head movement independently as an action unit (AU), with seven levels of intensities. FACS makes it possible to

understand how human facial and neck muscles generate different subtle expressions and movements. AUs are numbered arbitrarily [AU1 - AU66], e.g., AU1 corresponds to the inner brow raiser muscle.

- *FACS-based Action Units*. Based on the counseling literature pointing to the need for FACS-based facial expressions, we designed VHAs able to procedurally render the 65 facial action units using head and neck movements identified in FACS [47], with their seven intensities. In our design, blendshapes were created for each individual FACS AUs, which can be combined to create a large number of expression animations that can be rendered on the web. Because the VHAs facial animations were developed with custom-made blendshapes for each AU, the resulting animations allows for fine-grained procedural control over the VH's facial actions. The realism for the AU animations portrayed by each of our VHAs were evaluated by two FACS-certified experts. Their evaluation was conducted using a website created to display the agent in a neutral background and allow the activation of a single AU at a time. FACS experts were asked to identify which AU the VHA appeared to be portraying, rate the quality of the animation on a 5-point Likert scale (from "very poor" to "very well"), and to provide open-ended feedback. Based on the comments and scores received, the agents' AUs were iteratively improved until validated by the experts to ensure realism relative to a human portraying the same AU. The ability to specify which AUs are activated with what intensity to generate a VHA's facial or head animations facilitates systematic accumulation of knowledge, and future designs of VHAs' conversational cues.
- *FACS-based Facial Expressions*. VHAs should be perceived as non-judgmental, yet facially expressive. Toward that end we used three facial expressions found most relevant in counseling, according to the literature discussed above: *neutral*, *smile*, and *concern* expressions (a variation of sadness), shown in Fig. 5 and described in Table 2. The display of these expressions should be context-dependent. For example, if the user's answers about their drinking behaviors indicate that they are at-risk of severe consequences, smiling could be perceived as validation of the risky behaviors. In our current intervention, the VHA therefore displays a concerned expression, rather than a harsher negative expression, which would not support the non-judgmental counselor metaphor we aimed for. Furthermore, since counseling does not advocate for strong negative emotions, we activated mild concerned expression facial expressions at 50% intensity (cf. Fig. 5). Because MI counseling literature does not advocate for extreme emotions [121], we only activated positive facial expressions at 70% intensity. We discuss how we combined multiple facial expressions into a context-based multimodal model of active listening behaviors in Section 5.1.4.

Table 2. FACS-based description of Action Units used to generate facial expressions shown in Figure 5 and head movements for virtual agent counselors.

Counselor's Facial Expressions	Activated Action Units (AUs)	AU Intensities
Blinks	45 (blink) (15-20/min. randomly)	100%
Neutral	none	none
Smile	6 (cheek raiser), 12 (lip corner puller)	70%
Concern	1 (inner brow raiser), 2 (outer brow raiser) 4 (brow lowerer), 5 (upper lid raiser) , 7 (lid tightener)	50%
Nod	54 (head down), 64 (eyes down)	100%
Positive nod	smile & nod	70% & 100%
Negative nod	concern & nod	50% & 100%

- Blinks. Given that 3D VHAs simulate conversations in realtime, the VHAs should *blink* using Action Unit 45, on an ongoing basis at the random rate of 15-20 blinks per minute, similarly to humans.

Head movement Social Cue. We also used *head nod* animations as in previous literature for VHAs to establish rapport and working alliance [16, 58].

We included head nods in VHA design that are critical for maintaining rapport and for empathic active listening in counseling [3]. We designed three different head nods for context-dependent communication, by combining them with different facial expressions. These different nods are displayed based on the dialog current context, e.g. using an at-risk score calculated from the user's verbal responses to the VHA questions: different facial expressions nuance the message of the nod based on context (see Tables 2 and 3). We discuss how we combined multiple single social cues into a multimodal model of active listening behaviors in Section 5.1.4.

Eye movement Social Cue. The eye movement cue is defined as *the agent moves its eyes to intentionally or unintentionally fixate or track objects*. A study found that more time spent looking at medical chart lead to less patient satisfaction, whereas greater percent of time the clinician looks at patient was associated with greater patient satisfaction [70]. Our VHAs blink at a realistic rate (see Table 2), gaze toward the user when greeting, and maintain eye contact.

Arm and hand gesture Social Cue. The arm and hand gesture cue is defined as *the agent moves its arm or hand*. To address the prior limitations of VHAs which mostly modeled nonverbal cues in the categories of facial expressions, eye gaze, and nodding head movements (see Section 2), we created VHAs with fully rigged bodies. Our VHA body rig enabled us to create animations for arm and hand gestures relevant to counseling, including four single social cues: *resting*, *formless flick*, *2-handed flip*, and *2-handed contrast*, identified in Table 3, and shown in Figure 6. These hand gestures were then combined with other social cues to create a multimodal model of active listening social cues, discussed in Section 5.1.4, and shown in Figure 6 on page 26. These cues are added to the taxonomy in Table 10.

Additional new Kinesics social cues. As discussed in more details in Section 10, in our extension of the taxonomy we renamed or added the following Kinesics social cues, or values they can assume, to the original taxonomy because based on our literature survey, they are necessary for therapeutic communication, and we implemented them:

- **Viseme movement Social Cue:** we add the Viseme movement social cue that we defined as *the agent moves its mouth and face to match spoken words*, e.g., lip-synchronization during speech production. We map visemes to phonemes of the English language, to achieve realistic lip synchronization. This is an important social cue as a study revealed that users stop looking at a virtual agent that did not have lip synchronization which was found too disturbing for its users to continue looking at it during conversations [194].
- **Leg movement Social Cue:** our agent can stand up from its sitting position, walk to greet the user to respect social norms [85], and sit-down (see Figure 6 and Section 5.1.3). So we added a Leg movement social cue that we defined as *the agent moves its legs* with three single social cue values: *sit-to-stand*, *walk*, and *stand-to-sit*;
- **Posture shift Social Cue:** the Posture shift cue defined as *the agent moves its upper or lower body* include the social cues values we added: *twist*, *breathe*, *lean forward*, and *lean back movements*. Our agents simulate breathing with subtle up and down movements of the torso, can twist their torso before sitting down, lean forward to portray attentiveness and back.

5.1.3 Design Features for Visual Proxemics Social Cues. According to Feine et al.' taxonomy we use to describe our virtual agent [51], visual Proxemics cues refer to *the role of space, distance, and territory in communication*. It

includes two subcategories of cues: background and conversational distance. The proxemics social cue subcategory was found to be largely underrepresented. As we explain below, the background or environment in which an agent evolves represents an important social cue that can be designed to provide a therapeutic environment where users feel comforted and safe to disclose personal health information.

In what follows, we describe each of the visual Proxemics cues that we implemented in our agent, using Feine et al.' [51] taxonomy terminology and definitions, and literature-informed rationale for our design choices.

Background/Environment Social Cue. The background is defined as *the picture, scene, or design around the agent [that] forms a setting for the interaction.*

Although the taxonomy uses the term 'background' which can be a 2D background image, we renamed it to 'Environment' to account for the 3-dimensional nature of virtual agents' surroundings that they inhabit. As mentioned earlier in our discussion of the 2D/3D appearance social cue, 3D graphics tend to increase sense of presence (in addition to usability and positive UX).

The physical environment of the therapy room can indeed play an important role in clients feeling *comfortable* and able to *engage* [172]. A counselor's office should be a *healing* place which serves two functions [1]: (1) establish the role and prestige of the therapist in the client's mind in their expectations of the help to be received; and (2) provide a *safe* and *comfortable* place. Appearance and design include the layout of the room, which should convey expectations about what will happen in the room. Rooms used for therapy or counseling, therefore, should not use beds traditionally found in general physicians' offices or hospitals, but rather provide a '*homely*' *feel* [125]. Rooms with a clinical appearance were described as unhelpful.

"Soft rooms" - with plants, art posters, a decorative rug, and a padded armchair - were found to increase *disclosure* and to be perceived as *socially attractive* [125]. Indoor plants are considered *stress-reducing*, and can improve tolerance to pain [139]. Views of nature or a garden was also seen as beneficial to health outcomes [146, 163, 184]. Order and neatness appeared to affect the patients experience and their perception of therapist's *expertness*, *trustworthiness*, and *social attractiveness* [125]. The visual appearance of diplomas and awards in the counselor's office had a major impact on the client's initial perception of the counselor's *credibility* and *expertness* [71], which in turn influences client behaviors and attitudes [1]. Other features that emerged from the surveyed literature were preferences for a larger work space; warm, welcoming, non-formal, and *aesthetically pleasing* decor; natural light; and ample options for clients to sit [139].

Furthermore, the I-PEFiC model predicts that aesthetics of the agent embodiment is perceived by users and significantly impact user engagement, and satisfaction. We therefore extended the aesthetics design concept to the design of the 3D virtual agent's environment as well. Since VHAs aim to elicit users' socio-emotional responses similar to those just discussed about a counselor's office, we included most of the objects identified above. We also added bookcases with books to support the metaphor of a counseling expert, a vase with flowers on a credenza to add warmth as a natural element. The resulting office is based on the counselor's "soft room" office metaphor and is shown in Figures 4, and 6. The complete list of furniture and objects we included is also listed in Table 14 on page 47.

Conversational Distance Social Cue. Conversational distance is defined as *the spatial separation between the agent and its environment during a conversation.*

Interpersonal distance is a type of non-verbal communication which affects the relationships, including between patients and clinicians. Interactions typically occur in **personal sphere (1.5-4ft)** or **the social sphere (4-12ft)**. While optimal conversational distance is context- and culture-dependent [174], Hall's foundational work [60-62] established the "personal zone" to be typically reserved for friends and close acquaintances. In the context of healthcare and therapy, studies demonstrate that close distances can amplify interpersonal liking and facilitate collaborative performance (except in some special cases such as Eye Movement Desensitization and Reprocessing (EMDR) requiring side positioning for safety). Research indicates that physicians spending more time in the

personal sphere within three feet of the patient (i.e. 1.4-4ft) led elderly patients to comprehend instructions better [70].

The I-PEFiC model predicts that user engagement affects satisfaction with two main users responses: involvement (e.g. affective approach, rapport) and distance (e.g. affective avoidance, boredom). Hence, in an effort to increase engagement with positive involvement, we designed our VHAs to deliver the intervention sitting in front the user at eye level, at the equivalent of the personal sphere, as shown in Figures 4 and 6. The VHAs adopt a seated posture matching the user's position to maintain consistent eye level and interpersonal orientation throughout the interaction. Because research provides evidence that greeting contributes to establishing rapport and respecting social norms [85], we added a greeting scenario in which the conversational distance cue varies from the social sphere to the personal sphere, by combining kinesics and proxemics social cues, as described below.

5.1.4 Combining Verbal, Kinesics and Proxemics Social Cues to Model Multimodal Active Listening and Greeting Behaviors. Given the importance of active listening verbal and nonverbal behaviors in counseling, and of greeting scenario rituals (in most social interactions), we combined a variety of verbal, kinesics and proxemics social cues to create a multimodal model of communicative acts specifically relevant in the context of counseling and healthcare.

Active Listening. Our virtual counseling agent implements a coordinated system of rule-based empathic responses and active listening behaviors that operates through structured mappings between user's verbal responses, and both verbal and nonverbal therapeutic agent's communication. We aimed at modeling human counselors' interplay of nonverbal behaviors coordinated with the verbal behaviors discussed in Sections 4.1 and 4.2. Rather than attempting to infer emotional states through agent's interpretation of users' facial expressions which can raise privacy issues for the general public, or automatic emotion recognition which to date do not sufficiently take account the context of the interaction to be reliable [57], our agent employs contextually appropriate empathic responses, coordinated with systematic nonverbal behaviors that align with therapeutic communication principles.

The system's active listening capabilities are operationalized through five coordinated speech acts: **greeting, asking, listening, explaining, and farewell**, each mapped to specific nonverbal behaviors as detailed in Table 3, and shown in visual example is shown in Figure 6.

In the *asking mode*, the agent uses a single-hand formless flick gesture synchronized with its spoken utterances. In the *listening mode*, the agent leans forward to demonstrate attentive engagement and employs three types of expressive nods (neutral, positive, concerned) based on the user's response content and inferences made by the agent, sometimes followed by minimal acknowledgements (e.g. 'OK', 'I see'). In the *explanation mode*, the agent employs two-handed flip and contrast gestures to emphasize therapeutic content, while single-hand formless flick gestures in the asking mode create appropriate psychological space for user reflection and response. More details on the verbal content of the interaction are discussed in Section 5.4. At all times, the VHAs maintains mutual eye gaze, direct body orientation, uncrossed arms and legs, ongoing breathing movements with slight up-down movements of the torso, a blinking rate of 15-20 blinks per minute randomly, and lip synchronization based on phoneme-viseme matching.

This evidence-based approach ensures that nonverbal behaviors are grounded in actual therapeutic practice rather than computational assumptions about appropriate behavior. The agent's empathic responses utilize predetermined mappings between user responses about drinking patterns and clinically informed empathic language, while synchronized nonverbal behaviors reinforce the therapeutic communication intent. Each speech act serves distinct therapeutic functions: greeting establishes welcoming and respectful environment, asking creates space for disclosure, listening demonstrates engagement, explaining conveys knowledge with competence, and farewell provides positive reinforcement for completion, and adheres to social norms.

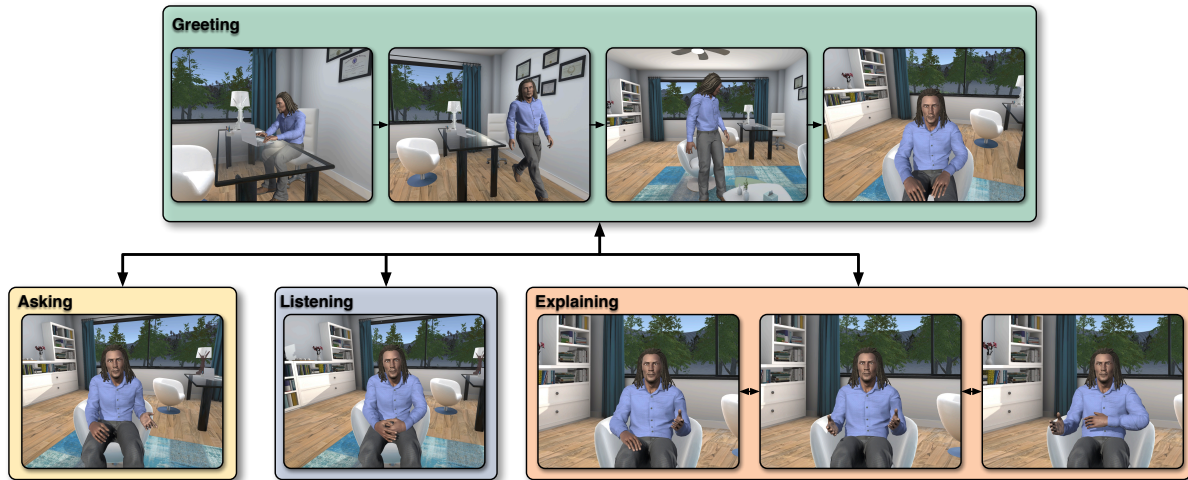


Fig. 6. Visual example of virtual agent counselor’s model of nonverbal communication: Top shows greeting scenario varying social distance (see Sec. 5.1.3), Bottom shows active listening communication combining speech acts with gestures (see Sec. 5.1.2). Details of communicative acts are shown in Table 3.

While this approach lacks the dynamic emotional inference capabilities of systems like ODVIC [98] or SimSensei [39], it offers the advantage of clinical reliability through evidence-based therapeutic communication protocols derived from actual counselor-client interactions. The coordination between verbal empathic responses and nonverbal active listening behaviors creates a multimodal therapeutic interaction that is both contextually appropriate and therapeutically sound, prioritizing clinical appropriateness over computational sophistication in emotion recognition and behavioral generation, that may introduce insufficient contextual sensitivity for meaningful social interaction [57], and privacy concerns.

Greeting. Furthermore, research has shown that greeting serve as important social rituals (according to culturally dependent social norms) that activate feelings of connection and warmth, contributing to positive feelings beyond what can be explained by personality factors or demographic variables [?]. In Western cultures, Kendon discussed the importance of greeting [85], and describes it as a sequential series of steps varying conversational distance. We therefore added a greeting scenario to start the interaction. The greeting provides a more “real” experience one might encounter in a real therapy session where the counselor will get up and greet you and show you to a seat. We therefore created animations for greeting using Kendon’s sequential steps: sighting and decision to greet, distant salutation, approach, close salutation, and lastly commencement of the conversation.

As shown in Fig. 6 on page 26, we simulated the counselor’s greeting with varying interpersonal distances: at the start of the session, the user first sees the large office space as if “standing at the office door,” and the VHA is *sitting* at its desk on an office chair in the *social sphere* (4-12ft), while *typing* on its computer without looking at the user, unaware that the user “is standing” at the door. The VHA then immediately *gazes* toward the user to convey that it has noticed that the user has “arrived” (i.e., has started the interactive session). The VHA then *gets up*, *walks* to one of the padded chairs, and *sits down*, after rotating slightly to check the chair location before sitting. We then vary the *camera angle* and lower it to simulate the user’s sitting down as well, so that the VHA is at *eye level* with the user, within the *personal sphere* (1-4ft).

Speech Act	Nonverbal Behavior Description	Therapeutic Rationale
Greeting	Types on computer, gazes toward user at the door, walks to padded chair in personal sphere distance, sits at eye level, uncrossed legs, hands on thighs	Establishes welcoming environment, signals transition from professional to therapeutic space, creates equal status positioning
Asking	Poses question, uses single-hand formless flick gesture synchronized with speech, leans backward slightly	Open gesture invites response, backward lean creates psychological space for user reflection and disclosure
Listening	Waits for user response; leans forward, maintains posture, nods expressively (three types: neutral, positive, concerned based on interaction current context), returns upright after answer	Forward lean signals active attention and engagement, varied nodding provides contextual feedback, sustained posture demonstrates commitment to listening
Explaining	Uses two-handed flip and contrast gestures synchronized with speech, maintains upright posture	Bilateral gestures emphasize key therapeutic content, upright posture conveys competence and knowledge
Goodbye	Congratulates for completing intervention while smiling, says goodbye while seated, maintains open posture	Positive reinforcement for engagement, seated position maintains therapeutic relationship rather than dismissive standing

Table 3. Coordinated Multimodal Verbal and Nonverbal Behaviors in virtual agent Counselor Communicative Acts, Lip-synching is performed for all spoken utterances.

5.1.5 Design Features for Visual Computer Mediated Communication (CMC) Cues. According to the taxonomy, the CMC cues refer to *visual elements that can augment or modify the meaning of a text-based message*.

This subcategory includes two low-level single social cues: *emoticons* (e.g. ;-), ^_^) and *typefaces* (design of letters and symbols). As discussed by authors [51], and evidenced by virtual agent literature, social cue subcategories such as CMC cues are very underrepresented, with virtual agent articles focussing primarily on features specific to the agent.

To ensure high usability, our VHA is embedded in a user interface described in Section 4, which includes **typefaces** to emphasize messages, and **graphical visualizations** to help users understand concepts discussed verbally the intervention (e.g., (e.g. bar graphs of Body Mass Index) (cf. Figure 3 on page 13, and Table 15 on page 48). In addition, we added **traditional user interface elements** to the taxonomy as they can help users understand the agent’s messages, e.g. we used a microphone on/off icon to facilitate turn-taking and signify to the user when the agent is listening (mic on) and when it not (mic off). We did not use emoticons.

5.2 Design Features for Auditory Social Cues

Auditory cues refer to *cues that can be heard (except words themselves)*. In the following, we describe each of the single Auditory cue that we implemented in our agent, using Feine et al.’ [51] taxonomy terminology and definitions, and literature-informed rationale for our design choices.

Auditory cues include two subcategories: (1) four voice qualities cues, including gender of the voice, pitch range, voice tempo, and volume; and (2) four vocalizations including grunts and moans, laughing, vocal segregates, yawn).

Voice quality. We used four commercial synthesized voices from *ReadSpeaker*¹ with different **gender**: two female voices (Julie, Ashley) and two male voices (Paul, James), which we respectively matched with female- and male- looking characters. We chose ReadSpeaker voices after a pilot study revealed that ReadSpeaker voices at the time of our study were perceived as less robotic and more varied than the widely used Ceroproc voices² voices.

The VHAs auditory cues were the least developed in this version of the VHA: we did not made any adjustments to the voice tempo and volume, nor did we express any vocalizations.

5.3 Design Features for Invisible Social Cues: Chronemics

Invisible cues refer to *cues that cannot be seen or heard*. Invisible social cues are separated into two subcategories: chronemics (i.e. first turn, response time) and haptics (tactile touch and temperature) as seen in Table 16. We did not include haptics social cues in our design.

First turn Social Cue. The first turn is defined as when *the agent starts the conversation and is the first to say something about a topic*.

In our system, the VHA was programmed to always starts an interaction. This approach was to put the conversant at ease and welcome them to the counseling scenario.

Response Time Social Cue. Response time is defined as *the amount of time it takes for the agent to respond to the user's input*, e.g. immediate response (0 second delay), slow response (2 second delay).

To implement and document this social cues, we conducted a quantitative objective evaluation of the VHA response time. We deployed the system before we conducted the study on a custom built server running Ubuntu 14.04 LTS with a Quad- core 2GHz CPU and 4GB RAM. We then measured the average response times of critical components of the system on 4G/Broadband connections between the two continents: North America and Europe.

The results are summarized in Table 4, showing very fast response times. More details on this objective evaluation are available in [19].

5.4 Design Features for Verbal Social Cues

In the following subsections, we describe each of the single Verbal social cue that we implemented in our agent, using Feine et al.' [51] taxonomy terminology and definitions, and literature-informed rationale for our design choices.

5.4.1 Content Social Cue. The content cues refer to *the strict and literal meaning of a message (i.e., what is said)*.

The taxonomy identified twelve low-level single social cues regarding content. The verbal cues from the taxonomy that are present in the therapeutic brief motivational intervention for alcohol use disorder and the additions we made to facilitate its delivery by a VHA were discussed in Sections 4.1 and 4.2. Sample excerpts of the VHA-participant dialog are shown in Table 1.

¹<https://www.readspeaker.com>.

²<https://app.cereproc.com/>

Table 4. Average response time and standard deviation analysis for VHA in milliseconds; 4G/Broadband connections over the Internet between North America and Europe; caching disabled (first run).

<i>Unity 3D character and office loading time at start of session</i>		
4G mobile data	30018.00 ms	+/- 663.75
Broadband internet	24626.50 ms	+/- 1910.78
<i>Text to speech (per sentence)</i>		
4G mobile data	939.38 ms	+/- 381.63
Broadband internet	551.07 ms	+/- 141.05
<i>Text to speech (per word)</i>		
4G mobile data	72.44 ms	+/- 40.05
Broadband internet	44.11 ms	+/- 23.42
<i>Speech recognition</i>		
In-browser (Chrome)	~30 ms	(Offline processing)

More specifically, throughout the intervention, principles common to effective BMIs known as *FRAMES* [122] are applied: Feedback about client's individual status is personalized (based on user's answers); Responsibility for changing is left with the individual (no persuasion to change is used, intrinsic motivation is nurtured); Advice is provided in a supportive manner; Menus of different options for changing that respect individual's readiness to change are offered; Empathic style of communication is central to the individual-clinician relationship (e.g. provides empathic feedback that does not raise resistance as shown in Table 1); and Self-efficacy is emphasized (with positive, encouraging messages).

From the *FRAMES* principles, we implemented the following 7 of the total 12 verbal cues identified in the taxonomy that are present in the original DCU content: **ask to start/pursue dialog**; **greetings and farewells**; **praise** to promote self-efficacy (e.g. "It's a good decision to look at your drinking Joe."); **refer to past** (when a session is resumed after interruption); **self-disclosure** (in the agent's introduction "My name is [X], a virtual agent designed to... I do not have full human intelligence but I can..."); **thanking** (e.g. "Thank you for sharing this."); and **(tips and) advice**. We did not include the following cues from the taxonomy: apology/excuse, joke, opinion conformity; self-focused questions, nor small talk (albeit some of these might be useful for VHAs).

While Feine et al.'s (2019) taxonomy provides a comprehensive foundation for general-purpose conversational agents, therapeutic applications require domain-specific content types. We extend the taxonomy with five content cues grounded in motivational interviewing and brief motivational intervention *FRAMES* principles [121, 122], which we implemented in our system:

- **Acknowledgment Social Cue:** defined as *the agent gives brief affirmative responses that signal attention and understanding without taking the conversational floor*. We derived the following acknowledgements based on the verbal codes generated by the clinical expert, also evidenced in counseling literature [127], combined with facial expressions and different types of nods shown in Figure 5: "Ah," "Alright," "Okay," "Good," "Sounds Good," "Oh, okay," "Great! Thanks for letting me know," "Oh, I see," "Okay, thanks!" These acknowledgments are distinguished from opinion conformity (where the agent takes a substantive stance on the topic) by their minimal semantic content and function in maintaining the user's conversational turn while signaling agent's active engagement.
- **Ask Social Cue** (to assess): defined as *the agent asks questions*. Feine's taxonomy only accounts for 'ask to start/pursue dialog', but MI requires asking questions for systematic assessment of behavior, readiness to change, and risk factors.
- **Explain Social Cue:** defined as *the agent providing information/education*, e.g. "a standard drink is...". Such psychoeducational explanations provide factual information distinct from advice or tips, and health behavior change interventions require clear explanations of concepts.
- **Feedback Social Cue:** defined as *the agent provides personalized feedback questions* based on user responses, which is a core *FRAMES* principle. This differs from advice-giving (already in Feine's taxonomy) by providing personalized normative information about user's status rather than prescriptive recommendations.
- **Refer to emotion Social Cue:** defined as *the agent acknowledges what user might feel*, e.g. when receiving normative feedback about their drinking that surprises or alarms them, e.g. "It can be surprising and (for some) discouraging to see that they fall higher on the alcohol use scale than they expected". Emotional validation and acknowledgment are fundamental to empathic counseling communication [156]. MI explicitly incorporates emotional reflection as distinct from content reflection.

Sample excerpts of the VHA-participant dialog are shown in Table 1, and intervention content is provided in Sections 4.1 and 4.2. Details of specific verbal content cues we used from the taxonomy are summarized in our discussion and shown in Table 16.

5.4.2 Style Social Cue. The style cues refer to *the meaningful deployment of language variation in a message (i.e., how something is said)*.

In the taxonomy, the style social cue includes five subcategories of individual social cues: abbreviations, formality, lexical diversity, sentence complexity, and strength of language.

We used **abbreviations** such as acronyms used in the DCU (with their explained meaning), e.g. BMI in graphs for Body Mass Index. **Formality** refers to *expressing words and sentences that are either compliant or in compliant with conventional rules*, e.g. the agent addresses user with last name instead of first name. In our implementation we complied to conventional rules, and the interaction is friendly, yet formal. One informality we introduced for our target users living in the USA, was for the VHA to address the user with their first name (which we asked them to provide), as practiced in counseling in the USA, thereby compliant with conventional rules. **Lexical diversity** refers to *the agent expresses many different unique words*, e.g. uses various synonyms such as “also”, “further”, “likewise”, “besides”. The DCU intervention content that the agent delivers has high lexical diversity.

Sentence complexity refers to *the complexity of the agent’s sentences in terms of number of used words, grammatical quality, and length of chosen words*, e.g. well-constructed sentences with the use of sophisticated words, or grammatically wrong sentences with simple short words. The DCU delivers personalized educational health information written at the reading level of the general US population, which stands around 6th-8th grade. Hence the language is purposely not complex, yet grammatically correct.

Strength of language refers to *the messages of the agent range from strong and assertive statements to submissive and equivocal statements*, e.g. “You should definitely do this” versus “perhaps you should do this.” For therapeutic applications following MI principles, we implement the equivocal end of this spectrum through empathic communication style characterized by: tentative phrasing (“might,” “perhaps,” “it seems”) rather than directive language; user-centered framing (“to you,” “for you,” “in your experience”); validating language (“that makes sense,” “it is understandable,”); reflective softeners (“it sounds like,” “I wonder if”).

For example, rather than assertive feedback delivery “You drink more than 90% of men your age”, we use empathic phrasing: “It might be surprising to you that 10% of men your age drink as much as you do.” This approach aligns with the FRAMES principle of empathic communication while instantiating Feine’s ‘strength of language’ dimension for therapeutic contexts. Furthermore, given that the DCU was written based on FRAMES principles, which emphasizes the communication of responsibility to be with the participant (no use of persuasion, intrinsic motivation nurtured), the agent does not tell the user what to do, but rather raises intrinsic motivation with statements such as “It’s entirely up to you whether or not you want to change your drinking.” Hence the language strength is not strong given Feine’s taxonomy definition.

The verbal style cues according to the taxonomy categories are shown in Table 16.

6 Quantitative Evaluation

We conducted a quantitative evaluation of the VHA acceptability, usability and feasibility. In what follows we provide a summary of our findings that we will use to show how they support the qualitative evaluation we conducted in the current study, in order to gather participants’ impressions in terms of *why* they were satisfied (or not) with the VHA. Additional details about the quantitative evaluation were reported in [19].

6.1 Participants

Participants were recruited online to complete the BMI intervention with the VHA and answer questionnaires. We recruited English-speaking adults residing in the USA who reported heavy drinking episodes (defined as consuming 5 or more drinks in a single occasion at least once in the past year) and expressed interest in receiving support but were not currently enrolled in treatment services. Participants (N = 136) were recruited online for a larger randomized controlled trial evaluating the content of intervention and its impact on change. Eligibility criteria included: (1) engagement in heavy drinking as defined above, (2) absence of current alcohol use disorder treatment, (3) no medical conditions contraindicating alcohol use, and (4) no severe mental illness. While the

control condition experienced a text-based intervention with no agent to interact with, for this study, we only included participants assigned to the experimental condition ($n = 51$) because they were required to have direct interaction experience with our virtual health counselors and therefore able to evaluate it.

The sample ($n = 51$) consisted of 62% male ($n = 32$) and 32% female ($n = 19$) participants, and one participant declined to answer; with a mean age of 28 years ($SD = 15.8$). Regarding racial identity, participants self-identified as White (42%, $n = 21$), Black (24%, $n = 12$), Asian (7%, $n = 3$), or Other (3%, $n = 2$), with 24% ($n = 12$) not providing a response. Most participants identified as non-Hispanic (84%, $n = 43$), while 15% ($n = 8$) identified as Hispanic and 1% ($n = 2$) did not respond. Nearly half of participants were married (46%, $n = 23$), with others reporting single status (34%, $n = 17$), divorced (7%, $n = 3$), or widowed (1%, $n = 1$); 12% ($n = 6$) did not provide marital status information. Educational attainment varied, with the largest group holding Bachelor's degrees (36%, $n = 18$), followed by some college experience (25%, $n = 13$), Associate's degrees (19%, $n = 10$), high school completion (14%, $n = 7$), and graduate degrees (6%, $n = 3$).

6.2 Procedures

Study participants were recruited via the internet. Targeted advertisements were placed on the Facebook (or Meta) social media platform, offering access to a brief self-help online program for alcohol users at no cost, and compensation for participation in the study. Individuals who clicked on the ad URL were redirected to a screening measure on Qualtrics to confirm study eligibility. Eligible participants were subsequently randomly assigned to receive an URL for either eEVA-4AUD delivered by the VHAs or a text-only version of the DCU interface. This study only includes those participants assigned to receive the eEVA-4AUD intervention. After random assignment, participants were given a username and password to access eEVA-4AUD. Participants had to log on to the website, complete a demographics survey, and activate the intervention.

6.3 Methods

Following intervention completion, participants provided feedback on the technology's feasibility, acceptability, and utility through a custom-developed questionnaire. Although studies that aim to establish causality between the introduction of agent's feature(s) on users' perception or behavior using the experimental design method can provide insightful causal findings (e.g. comparing an agent with empathic cues to a baseline agent without such cues as in [98]) to claim that empathic cues cause an increase in user's intention to use the system), our current study aims to focus on constructing an accurate description of what is happening when we introduce some agent's social cues - using descriptive methods such as questionnaires (this section), as well as interviews (see Section 7). We plan to conduct experimental research on specific social cues based on our current descriptive findings, which is a well accepted ascendingly ordered behavioral research methodology in virtual agent research [157].

Given the absence of standardized instruments (and of benchmark) for evaluating interactions with VHAs of varying complexity levels, we adapted and combined elements from established questionnaires commonly employed in human-technology interaction research involving social cues, including both graphical embodiments and robotic systems. The questionnaire integrated components from three validated frameworks: O'Brien and Toms' engagement model [130], Heerink et al.'s Almere model [67], and Bartneck et al.'s "Godspeed questionnaire" [9].

O'Brien and Toms' engagement model [130] assessed VHA acceptability through **engagement** constructs, examining six technology attributes that enhance user engagement likelihood: challenge, interest, motivation for use, and technological appeal (e.g., "I found the health assistant interesting"). The 7-point Likert scale maintained consistent scoring (1=Strongly agree; 7=Strongly disagree).

The Almere model [67] evaluated digital health agent **acceptance** using constructs derived from the Unified Theory of Acceptance and Use of Technology [188]. Participants responded to 13 statements (e.g., "I enjoyed participating in this session with the health assistant") corresponding to constructs that predict technology usage intention and perceived utility. The same 7-point Likert scale format was employed (1=Strongly agree; 7=Strongly disagree).

The Godspeed instrument has been extensively utilized in evaluation of human-like characteristics in human interactions with socially interactive agent (virtual or robotic), measuring five core constructs—anthropomorphism, animacy, likability, perceived intelligence, and safety—that have proven valuable for assessing interactions with both virtual agents and robotic systems. The Godspeed questionnaire [9] is comprised 12 items measuring five constructs that assess **human-like characteristics** in robots, which we adapted for VHA evaluation by replacing 'robot' with 'health agent' (e.g., "The health agent seemed warm"). The constructs included: (1) anthropomorphism (ranging from rigid to elegant movement), (2) animacy (mechanical versus organic qualities), (3) likeability (unfriendly to friendly spectrum), (4) perceived intelligence (incompetent to competent range), and (5) perceived safety (anxious to calm continuum). All items utilized 7-point Likert scales from (1=Strongly Agree; 7=Strongly Disagree), where lower scores represented more favorable outcomes in terms of humanlikeness. No threshold criteria were established as the measures served descriptive purposes within this study's framework. Each item included an optional "Comments" field for participants to provide optional qualitative elaboration on their numerical ratings.

For analysis, agreement was operationalized as selections of somewhat agree, agree, or strongly agree. Reported values represent the percentage of respondents selecting any of these agreement categories.

6.4 Results

Participants reported high levels of **acceptability** and **perceived utility** of the VHA, as indicated by responses to the Almere model items (mean = 2.31, SD = 1.05). A large majority (83.8%) agreed—i.e. somewhat agree, agree, or strongly agree—that they enjoyed interacting with the agent, describing the interaction as novel and reflective. Most participants found the agent physically appealing (71.6%) and appreciated its pleasant voice (79.7%). Results are shown in Table 5.

The agent was also found to be **easy to use**: 86.5% of participants reported that they quickly learned how to interact with the system and found it intuitive. However, some experienced difficulty with the voice recognition feature, with only 67.6% feeling that the agent understood them when speaking—an issue often attributed to user-side microphone settings. Results are shown in Table 6.

Despite this, 74.3% felt that they could have a meaningful conversation with the agent, and 68.9% reported that it sometimes felt like talking to a real person. The agent was widely perceived as friendly (86.5%) and helpful (85.1%), and 90.5% of participants indicated they were comfortable disclosing sensitive information to the agent. Notably, 83% reported greater comfort disclosing their alcohol use to the agent than to a physician. Trust in the agent's advice was high (82.4%), and 85.1% intended to act on it.

Question	Answer
Enjoyed interacting with agent	83.8%
Agent was physically appealing	71.6%
Agent had a pleasant voice	79.7%
Felt like talking to a real person	68.9%
Agent was friendly	86.5%
Comfortable disclosing information	90.5%
More comfortable disclosing sensitive information to agent than to physician	83.0%
Trust the agent's advice	82.4%

Table 5. Participant Endorsements of Agent Acceptability

Question	Answer
Easy to interact with the health agent / Learned to use quickly	86.5%
Agent understood me	67.6%
Can have a conversation with agent	74.3%
Agent was useful	85.1%
Plan to implement agent's advice	85.1%

Table 6. Participant Endorsements of Agent Usability and Advice

In terms of **engagement**, participants scored highly (mean = 2.86, SD = -0.96), with 69% feeling confident using the technology. The system was perceived as practical and adaptable, with 90% agreeing that it could support other health concerns. Additionally, 88% found the agent interesting, and 85% expressed willingness to interact with it again. Results are shown in Table 7.

Assessments of **human-likeness**, based on the Godspeed questionnaire (mean = 2.07, SD = 0.89), further reflected positive impressions. Participants endorsed a range of human-like qualities, including responsiveness (89%), warmth (90%), knowledgeable (92%), and respectfulness (90%). Results are shown in Table 8.

Question	Answer
Not worried about making mistakes	69%
Good idea to use agent	86%
System can be adapted to my needs	90%
Agent is interesting	88%
Would interact with agent again	85%

Table 7. Participant Endorsements of Agent Value, Adaptability, and Reuse

Therapeutic Personal Human-like Characteristics	Answer
Moves appropriately	85%
Seems warm	90%
Responsive	89%
Knowledgeable	92%
Flexible	83%
Honest	90%
Respectful	90%
Relaxed	90%
Confident	92%
Interested	86%
Open-minded or nonjudgmental	84%
Supportive	89%

Table 8. Participant Endorsements of Therapeutic Personal Characteristics

Overall, findings suggest that the digital health agent was perceived as **acceptable, engaging, and exhibiting strong therapeutic personal human-like characteristics**, indicating that VHAs designed with the features we implemented seem viable for delivering brief motivational interviewing-based interventions.

We also sought to investigate further and to identify exactly *which* VHAs' features users noticed, and *how* they felt about the VHAs' various social cues we designed, as we discuss in the next section.

7 Qualitative Data Gathering and Analysis for Intended Users

7.1 Research design rationale

We chose a qualitative approach for the current formative evaluation of the eEVA-4AUD prototype to (1) reach a deep understanding of our target end-users' perspective on the use of VHAs in their daily lives (difficult to capture from questionnaires and quantitative analysis [137]), (2) explore user perceptions and socio-emotional responses to VHA design features that have not been well studied, and (3) derive design considerations for future VHAs. Even though *qualitative methods* with interviewing are particularly well-suited to elicit users' impressions on design features that are novel, complex and intertwined [93, 95] (such as VHA features), recent reviews found that very few grounded-theory based qualitative studies have been conducted on VHAs [102, 180].

7.2 Participants

We asked all participants who completed the intervention (discussed in Section 6) if there were willing to participate in interviews. A subsample of study participants ($n = 20$) agreed to participate in this qualitative study after completing the VHA-delivered intervention.

Among the participants, 50% identified as men, 45% identified as women, and one participant declined to answer. Ages ranged from 22 to 52, with a mean age of 35.67 (SD = 8.04). Because we conducted this study in the USA, we used the categories used by USA Census Bureau at the time of our study to gather participants' racial and ethnic demographics. Participants reported their race and ethnicity demographics as follows: White (50%), African American (20%), Asian (5%), and Other (5%), with no American Indian or Alaska Native, and no

Native Hawaiian or Other Pacific Islander. Twenty percent of participants declined to answer. When considering the USA Census Bureau ethnicity, 70% identified as non-Hispanic, 10% identified as Hispanic (10%), and four participants declined to respond. Participants reported that they played an average of 2.5 hours of games per week, with a range of 1 to 5 hours per week.

7.3 Procedures

Following completion of the intervention, a principal investigator completed data quality checks by assessing quantitative responses to a reverse scored item. All participants with good quality quantitative responses were invited to engage in a qualitative interview to gather feedback about their experience with the VHAs from the eEVA-4AUD intervention across several dimensions. Twenty (20) participants agreed and were contacted individually by telephone and completed an interview. However, although twenty individuals agreed to participate in the present study, one recording was inaudible, and another recording failed. Therefore, a total of $n = 18$ recordings were used in this analysis. A recent systematic review of empirical tests indicated that, across many different approaches, saturation was reached within an average of 12–13 interviews, demonstrating strong external reliability [69]. We obtained saturation of responses with our 18 participants.

We scheduled the interviews as soon as the participants were done going through the intervention. Participants did not have access to the platform during the interview and referred to their memory of the agents to answer questions. The interviewer could follow-up with additional clarifying questions as needed. Interview audio was recorded by the facilitator, and transcribed by an external service. Transcriptions were reviewed, compared to the original interview audio, and corrected as needed by two graduate students.

7.4 Post-Intervention Semi-structured Interviews

We conducted semi-structured interviews with prepared open-ended questions addressing several dimensions of the interaction with eEVA-4AUD. Participants were asked about (1) their experience with the VHAs, e.g., VHAs' appearance and body language, voice (tone and speed), diverse physical characteristics, and potential addition of an adaptive social intelligence that could perceive and interpret users' social signals; (2) feedback regarding the virtual office environment and their opinions on personalizing the digital office and window view; (3) the intervention content of the eEVA-4DCU program such as impressions on each section, and the pace of the material; (4) their comfort level with the technology, and any barriers to using the eEVA-4AUD system effectively (e.g. bugs for programmers to fix in the next version). Although questions were asked in a specific order, there was no specified order or level of response required.

7.5 Data Analysis

We used *grounded theory* [55], and did not formulate a priori hypotheses. Instead, direct feedback from alcohol users of the VHA was gathered. We began inductively, using open and axial coding as detailed below. To ensure reliability and dependability, we utilized an iterative “coding by consensus,” and measured inter-rater reliability.

Open coding was first completed independently by two graduate students wherein each interview transcript was examined to identify recurring themes, key words, and phrases related to the intervention.

Axial coding was then performed, where these coders independently identified recurring words and phrases within the transcripts and subsequently organized them into hierarchical groupings [137]. These categories gleaned from axial coding were then used to create a codebook for the present study (provided in Appendix A).

Next, *selective coding* was used to select excerpts from all transcripts, and both original coders tested the codebook on these excerpts to confirm viability and suitability to accurately code the data. After the codebook was completed, two additional graduate students were trained on using the codebook and double coded all using

*Dedoose*³ qualitative coding software. After coders independently coded small batches of interview transcript data, they met with the principal investigators and an additional graduate student. During these meetings, the principal investigators and graduate students acted as additional coders to review all codes, resolve disagreements or inconsistencies, and determine the appropriate code(s) for excerpts where the two coders conflicted [137].

Inter-rater reliability was conducted. Codes that had a low baseline (were not present enough to run reliability) were excluded ($n = 1$ code). The “Other” codes (did not fit in any other coding category) were also excluded from inter-rater reliability analyses ($n = 4$ codes). As shown in Table 17, inter-rater reliability ranged from moderate agreement ($\kappa = 0.59$) to perfect agreement ($\kappa = 1.00$).

After coding was completed, *frequency counts* for all themes were computed. This allowed the most frequently experienced themes to be identified. Table 17 in Appendix A shows the details of our resulting Code Book, with theme codes and subcodes definitions, frequency counts for theme occurrences (n), and reliability estimates (i.e., Cohen’s kappa statistic) associated with each code. Positive and negative sample excerpts are included below and in our design summary tables (see Tables 9, 10, 11, 12, 13, and 14), to illustrate users’ responses to the various design features.

Codes were also *weighted* by coders who assigned ratings on a five-point Likert-type scale based on the valence of the users’ comments (1 = very negative comment, 2 = somewhat negative comment, 3 = neutral or mixed comment, 4 = somewhat positive comment, 5 = very positive comment).

8 Findings of Qualitative Analysis

8.1 Codes and Thematic Overview

We first began by identifying patterns in the data from our investigation. Reporting on qualitative studies is invariably lengthier than for other scientific methods, because it requires detailed descriptions of design rationale, and quotations from raw data excerpts to illustrate our findings, to “bring them to life” [95]. As a result, we cannot delve into detail on all of the themes identified in this manuscript, and instead we focus on the themes related to the VHAs and office’s social cues design, with the feelings they evoked in users.

A total of eighteen transcripts were coded from eighteen semi-structured interviews with individual alcohol users who engaged in the intervention. After open coding for thematic analysis, several main themes related to the VHA were identified as parent codes and corresponding sub-codes, including:

- **agent acceptability**, with corresponding subcodes such as general user experience, personal qualities, appearance, nonverbal behaviors, voice / verbal tone and speed, hypothetical sensing social intelligence;
- **virtual office environment**, with corresponding subcodes such as general user experience, realism, and preferences / view;
- **agent diversity**, with corresponding subcodes such as ethnicity / race, age, and gender.

Our qualitative analysis code book is provided in the Appendix in Table 17. In what follows, we discuss our findings, organized by the main themes: VHAs’ acceptability, office environment, and diversity, and each of their sub-themes.

We provide main *frequency counts* (n) of comments on each theme per transcript and across all transcripts, *weighted averages* (M) of weights assigned by coders on a 5-point Likert scale based on the valence of user’s comments (1 for very negative comment, 5 for very positive comment), and *key quotes* from the qualitative analyses. The weighted averages of the coders’ weights of the valence of users’ comments for each of the sub-themes discussed are shown Figure 7, sorted by decreasing weighted average.

Below, we present the themes that arose from the semi-structured interviews. The term “transcript” is used when referring to a full alcohol user’s interview. The term “excerpt” is used to describe a section of the interview

³<https://www.dedoose.com>

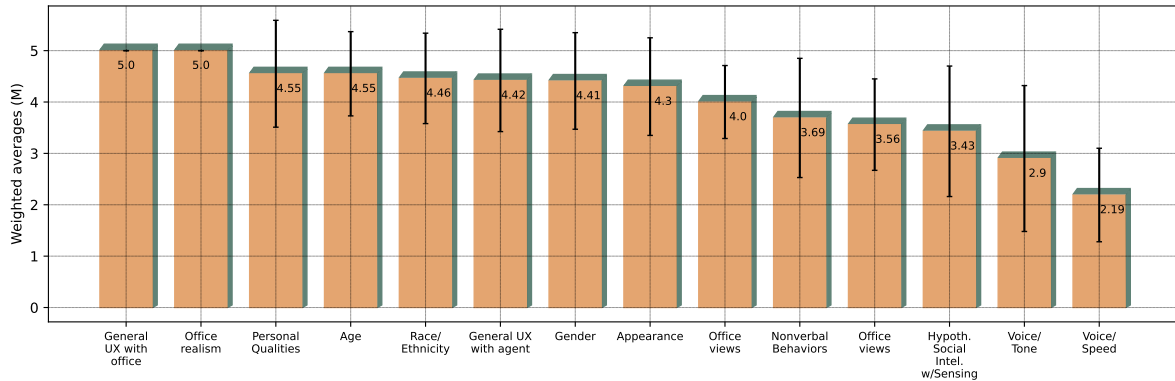


Fig. 7. Weighted averages of weights scored by coders based on the valence of the users' impressions of agents' individual social cues, using a five-point Likert-type scale (1 = very negative comment, 2 = somewhat negative comment, 3 = neutral or mixed comment, 4 = somewhat positive comment, 5 = very positive comment). UX stands for User Experience.

identified as a unit of analysis that applied to one or more codes. An excerpt can be a sentence or several sentences if the content was connected.

8.2 Acceptability of the virtual agent as Counselor as a whole

Themes surrounding the acceptability of the VHAs' visual and vocal features were mentioned 223 times across participant interviews (39.6% of total excerpts), and every individual mentioned at least one aspect of agent acceptability. Users commented on the VHA's *appearance* ($n = 24$, present in 83.3% of transcripts), *non-verbal behaviors* ($n = 27$, present in 83.3% of transcripts), *voice tone* ($n = 18$, present in 88.9% of transcripts), and *voice speed* ($n = 23$, present in 88.9% of transcripts).

Each subsection below discusses our findings for each identified acceptance of the sub-categories, with corresponding weighted values of scorers' ratings of users answers on a scale from 1-5 (5 = very positive) (see Fig. 7), and includes examples of user's impressions and feelings from a subset of excerpts.

8.2.1 Users' general user experience (UX) with the VHA. users' general UX with the agent ($n = 12$), i.e., personal feelings or emotional responses stemming from the interaction with the agent and/or experience with the agent selection menu, received very positive feedback ($M = 4.42$). One user noted:

"It was more comfortable after, I don't know, maybe 10 minutes. I forgot it was, you know, artificial intelligence. I feel like I was talking with a real person. But yet I didn't feel that uncomfortable like, you know, I would talking with someone, one on one."

As shown in Table 9, most users expressed that they felt *"very comfortable talking with the agent"* and that they felt they were *"talking to a real person."* Multiple users felt that the interaction with the agent was *"similar to therapy."*

8.2.2 Personal qualities of the VHA. users attributed personal qualities toward the VHA ($n = 11$). As shown in Table 9, although some users, felt that the VHA was *"truly just a piece of software,"* and that the VHAs had *"no emotions,"* most comments were rated very positively ($M = 4.55$), with users expressing that they felt that the agent was *"very caring," "very calm, and reassuring," "understanding," "informative," "impactful," "very concerned,"* and *"friendly."* One user mentioned:

"I thought that she was very concerned. That she uh she cared a lot and there's no negative things to say about it. I thought that she paid attention she took me through the, you know what I mean, without a problem."

8.2.3 **VHA's physical appearance.** The VHA's appearance ($n = 24$) was rated positively ($M = 4.3$) on multiple appearance design features: realism, attractiveness, clothing/style (see Section 5.1.1).

VHA's **realism** [appearance]. As shown in Table 10, a few users interpreted the VHA's graphical realism social cue as a signal that made them look *"computer generated," "like an anime character,"* with one user mentioning:

"Obviously it wasn't like, high definition.[...] Like it wasn't like a video game."

Most users, however, found that the VHAs looked realistic, e.g., *"like a regular person,"* Users felt that the VHAs were *"appropriate for the application,"* and that they *"really liked"* the VHAs. No user reported feeling uncomfortable and many users indicated they felt *very comfortable*. One user specified:

"She was friendly, she didn't creep me out or anything".

VHA's **attractiveness** [appearance]. Two users mentioned that VHAs' attractiveness cues we designed signaled that the VHAs were *"pretty attractive looking,"* and that their *"attractiveness definitely helps."* One user mentioned:

"Um, I think it definitely helps to have someone that you know, is not, I don't want to say it's like good looking, but just not, like not bad looking".

VHA's **clothing and style** [appearance]. As shown in Table 10, users did notice and comment on the VHAs' clothes and styles. They identified the clothing style cue as a social signal of *"casual like,"* and *"well dressed,"* which was our goal. One user said :

"I liked how he was well dressed. It seemed like his coat wasn't buttoned if I remember right. So he was well dressed and professional, but you know just a little bit relatable like laid back casual like, Hey, you can talk to me."

Another participant stated, *"Maybe, their appearance. [...] It just looked like a young professional."* Users detailed the VHAs' hair, and attributed the hair style to the VHA, which facilitated feelings of being *"attached to"* or *"connected with"* the VHA. As an example, one user mentioned:

"Uh, I, particularly liked the guy who had like, had this Rastafari hair, or somewhat Rastafari hair. [...] Yeah. I felt that it was the one whom I connected to most."

8.2.4 **VHA's active listening nonverbal social cues.** The VHA's nonverbal behaviors ($n = 27$) were received positively ($M = 3.69$) (cf. Table 11). A few users thought that the body language looked somewhat unrealistic, with one user noting:

"Yeah. [the body language] was the part that made you kind of realize, okay, this isn't real.",

another user mentioning that it was *"a little stiff"*

However, most users perceived the VHA's NVB social cues positively ($M = 3.69$), and as *"pretty pretty realistic, their motions and things,"* and as having *"a nice body language."* One user said:

"Yeah, I actually browse through a few of them. And they, they seem pretty, pretty realistic. Their like motions and things."

The VHA's body language social cues were perceived multiple times as signaling the VHA's openness, and triggered users' responses toward the VHA of being *"non judgmental"* and *"non confrontational."* The VHA's subtle active listening social cues were also perceived as *"little different gestures,"* which succeeded in eliciting the user's social response that the VHA was *"really attentive"* and *"listening"* to the user, and users *"liked"* the

VHA. The active listening NVBs coordinated with simple verbal reflection social cues were also perceived and triggered a sense of being “validated” by the VHA.

“I was really really impressed with the body language design. It was like very open body language”.

Users perceived the VHA’s **greeting** (see Section 5.1.2 and Figure 6 on page 26). One user said:

“And then I come in, it’s like coming into an appointment.”

Another user said:

[...] And then when I went back in several times, you know, like, he would be at his desk. and he would come over and be like, Hey, you know, I remember what we talked about last time and you don’t worry about that. so, Yeah, that was awesome. I thought it was really great.

8.2.5 **Hypothetical VHA’s sensing and adaptive social intelligence.** When asked about the potential addition of a sensing module for the interpretation of their facial expressions ($n = 7$), users had contrasting feelings ($M = 3.43$). One user said:

Oh, that would be really great and I think that’s that’s really important in it being able to read me too.

On the other hand, one user expressed concerns:

- Yeah, I don’t really feel like comfortable with like facial recognition stuff in general. Personally, I wouldn’t really want to do that.

Yet, another user suggested that there be an option added to enable it so that people do not get “nervous.”

8.2.6 **VHA’s verbal tone, and voice speed.**

VHA’s verbal tone. The VHA’s voice and verbal tone ($n = 18$) received very mixed results ($M = 2.9$). Some users found the voices “pleasant,” “pretty realistic,” or “natural.” One user perceived that the voice “was good, it matched what they look like”, and another one said:

I thought [the voice] was natural to me. Like being there with that person.”

One user mentioned:

“His voice sound urban. like he is- From my, from he’s probably black just like me. So that’s probably one of the reasons why I’m picking, he sounded like me.”

Others found the voices “too text-to-speechy,” “robotic,” “a little too artificial,” or “monotone.”

Many users felt the active listening minimal **acknowledgements** were “somewhat repetitive,” and some users described them as “a little annoying after a while.” One user, however, felt validated and said:

“So yeah, ... he would be like, “mm hmmm” like, “thanks for sharing that.” I don’t remember what he would say. so yeah body language, that kind of validation of what I was saying, and [it] feels good yeah.”

VHA’s voice speed. The voice speed ($n = 23$) received the worst rating of all the VHA design features ($M = 2.19$). Although one user mentioned that “it was perfect,” most users reported that they found the voice “too slow.” Multiple users suggested that an option to adjust the voice speed be provided, with one user mentioning:

“Yeah I would like to speed up to some preferred speed. You know, like where on Youtube there is an option where you can speed up a video, I would like that option.”.

8.3 3D Virtual Office Environment

The virtual **office environment** (see Figures 4 on page 14 and 6 on page 26) was brought up by two-thirds of participants, across 26 excerpts (4.6% of all excerpts). Users positively rated the realism of the virtual office setting ($M = 5$) and the view outside the office window ($M = 4$). As shown in Table 14, participants also reported positive feelings regarding their general experience with the virtual office ($M = 5$). Users perceived the environment cues

as signaling an office cue “*like sitting in an office environment*,” and a working space cue with the VHA “*is working on his stuff*” Users had the following socio-emotional responses to the office: it was “*really comforting*,” “*very comfortable*,” “*very therapeutic*,” and “*not bland nor overdone*.”

One user said:

I think it was good. Like it was really comforting.

Another user mentioned:

Like I felt like the room was a very good kind of a therapeutic setting type room.

One user was particularly pleased with the “*clean setting*” and said:

I thought they were uh in a clean setting. The setting was cool cuz I really liked, you know, to be in a nice setting myself. .

8.4 Agents’ Diversity

Participants were presented with a menu of 25 visually distinct VHAs to choose from, who portrayed cues contributing to the perception of race, age, and gender demographics (see Fig. 2 on page 12). As shown in Table 12, participants did perceive the VHAs’ gender, race, and age social cues that we designed as social signals by attributing gender, race, and age characteristics toward the VHAs. Participants commented on these VHA’s diversity social signals, which were mentioned 71 times (12.6% of all excerpts, noted by 100% of participants), and which were perceived positively. For example, one participant stated:

“I like that you can choose which you know which one you wanted, younger, older, man, woman, different races.”

All participants mentioned at least one aspect of the VHAs’ diversity social signals, including excerpts regarding perceived gender ($n = 18$, 25.4% of agent diversity excerpts, noted by 55.6% of participants), perceived race ($n = 14$, 23.9% of agent diversity excerpts, noted by 44.4% of participants), and perceived age ($n = 11$, 15.5% of agent diversity excerpts, noted by 33.3% of participants).

Several participants stated they chose a virtual health agent that was like them in at least one diversity characteristic. As one participant explained:

So, um, I’ve been to therapists too. And I think when I choose someone, I kind of want someone that’s similar gender and ethnicity for me. So that is why I chose someone that was like, also Asian and I think this one was Japanese.

A few users interpreted the social cues of the “*older male*” as VHA’s social signals, which triggered social responses that attributed *wisdom* and *experience* to the VHA, with one user mentioning:

“I just picked the older gentleman because it seemed like he would be more like a good therapist.”

9 Discussion

The findings presented in this study represent an advancement in the development of comprehensive virtual counseling systems that address critical gaps in mental healthcare accessibility and delivery, and in virtual agent research.

Context. The growing demand for accessible mental health services, combined with stigma barriers that prevent many individuals from seeking help for sensitive issues like alcohol misuse [31], has driven significant research into virtual health agents designed as counselor extenders: i.e. they can extend human counselors’ practice within the healthcare ecosystem with, for example, preventive interventions, just-in-time interventions 24/7, booster follow-up sessions. This need is particularly acute for stigmatized conditions, where individuals may be

reluctant to seek traditional counseling, creating opportunities for virtual agents to provide initial low-intensity interventions for stepped care intervention when people are too apprehensive to engage with human counselors.

Indeed, in December 2025, the US National Center for Health Workforce Analysis (NCHWA) released projections for 2023-2038 documenting the increase in demand for behavioral health, mental health, and substance use disorder services in the United States. Projections indicate persistent workforce gaps: by 2038, shortages are expected to reach approximately 77,050 addiction counselors, 99,780 mental health counselors, 99,840 psychologists, 43,810 psychiatrists, 33,840 marriage and family therapists, and 39,680 school counselors. As behavioral health needs continue to rise, rural counties are more likely than urban counties to lack behavioral health providers [84, 118, 145, 164]. The documented disparities in behavioral health access among racially and ethnically diverse populations [178], and their frequent treatment by non-Latino Caucasian providers [43], highlight the potential for diverse virtual counselors that can provide culturally representative mental health support to underserved communities.

Research Questions. This discussion addresses how our findings contribute to answering our research questions, while situating them within the broader virtual counseling literature (see Sections 3.3 and 5):

RQ1: *What design choices can inform the development of a virtual health agent as counselor that elicits similar socio-emotional responses as those experienced with an effective human counselor?*

RQ2: *What design choices can inform the development of a virtual office environment for a virtual agent as counselor that triggers similar socio-emotional responses as those experienced in an effective human counselor's physical office?*

RQ3: *What socio-emotional responses do diverse adults exhibit when they can choose from a diverse set of virtual health agents as counselors for a counseling intervention?*

Format of Findings Discussion for Designers of Virtual Agents as Counselor Extenders. While we have not yet established causal relationships between specific social cue designs and user socio-emotional responses, our findings suggest the effectiveness of our theory-based metaphorical approach. The convergence of positive quantitative outcomes (Section 6) with qualitative responses (Section 8) indicates that our design framework—incorporating 77 individual social cues within multimodal communicative acts and a therapeutic 'soft room' environment—successfully achieved two key objectives: (1) users perceived the VHAs as possessing evidence-based characteristics of effective counselors [2], and (2) users exhibited socio-emotional responses consistent with those observed in effective counseling interactions.

Our discussion below also includes Design Tables that map our implementation of individual and combined social cues to observed socio-emotional responses. These tables provide VHA designers with empirically-informed guidance on which cues may elicit specific user responses that influence system usage intentions and satisfaction, and consider including them in their design. Following our explanation about how a cue becomes a social cue when it triggers a socio-emotional response (see Section 3.2), for each category of themes we identified in our qualitative analysis, the Design Tables include the following information:

- (1) details about which VHA's social cue (or combination of) we implemented and how we did it, shown in Column 1;
- (2) the social signals associated with these cues perceived by users as revealed in the interviews, shown in Column 2;
- (3) the socio-emotional responses they triggered in users as revealed in the interviews, shown in Column 3;
- (4) additional sample excerpts of these socio-emotional responses from the interviews, shown in Column 4, with the weighted average, M , of the coders' scoring of the users responses on a 5-point Likert scale (1 = very negative comment - 5 very positive comment).

Table 9. Users' general user experience with virtual health counselors (VHAs) as a whole, i.e. including the **combination of all designed social cues**, and **VHAs' personal qualities** users perceived and encoded about the VHAs, with sample excerpts of interviews, and weighted averages of weights (M) assigned by coders on a 5-point Likert scale based on the valence of user's interview comments (1 for very negative comment, 5 for very positive comment, cf. Figure 7).

All VHA's social cues	All VHA's social signals	User's socio-emotional responses	Sample Excerpts of Users' socio-emotional responses to UX with VHA's (M=5.0) and VHA's PERSONAL CHARACTERISTICS (M=5.0)
General User Experience (M=5.0)	all of the VHA's social signals	comfortable with, forgot it was AI, like talking to someone one on one, like therapy	<ul style="list-style-type: none"> - <i>It was more comfortable after, I don't know, maybe 10 minutes. I forgot it was, you know, artificial intelligence. I feel like I was talking with a real person. But yet I didn't feel that uncomfortable like, you know, I would talking with someone, one on one.</i> - <i>Overall, I thought it was really interesting. Kind of like self-reflective, like therapy type thing. You know, this definitely helped me think about how I use alcohol on a day to day basis.</i>
Personal qualities (M=5.0)	all of the VHA's social signals	very attentive, cared a lot, calm, reassuring, understanding, informative, impactful, laid back, non judgemental, very concerned, friendly, comfortable with VHA, has positive impact	<ul style="list-style-type: none"> - <i>I thought the person too was very like, calm and reassuring, like, I thought it was a good bot, you know, I felt - it's weird to say I felt like I had an interaction with the, like, the format and, you know, the way they presented information and, overall, I think it impacted my mindset. Definitely.</i> - <i>It seemed, not judgmental, you know, understanding pretty laid back not not, you know, ready to judge you or you know... Analytical.</i> - <i>I thought that she was very concerned. That she uh she cared a lot and there's no negative things to say about it. I thought that she paid attention she took me through the, you know what I mean, without a problem.</i>

9.1 Evidence-Based Virtual Brief Motivational Interviewing Implementation (RQ1)

The adaptation of Brief Motivational Interviewing for virtual agent delivery represents a significant step toward standardizing evidence-based interventions in digital health platforms. Recent systematic reviews have identified significant gaps in the integration of MI principles within VHAs, with only 21% of studies covering all MI principles and 29% including all processes [116]. While the current implementation does not implement all MI principles either, it demonstrates successful integration of evidence-based BMI protocols, with 82.4% of participants trusting the advice given by the health agent and 85.1% planning to follow that advice.

Particularly compelling is qualitative evidence that participants recognized the agent's adherence to BMI principles. One participant noted, "I didn't feel like I was given advice, more like information to be able to make my own decision. I was the one with the power to give myself advice," well capturing the client-centered, autonomy-supporting approach fundamental to motivational interviewing. These finding suggests that virtual agents can successfully implement the collaborative stance essential to BMI effectiveness [175].

9.1.1 General User Experience with the virtual agent Counselor: Most participants felt that their overall user experience interacting with the VHA was very similar to having a therapy session and talking to a real person, yet users were not confused about interacting with a digital artifact. As one participant explained: "It was more comfortable after, I don't know, maybe 10 minutes. I forgot it was, you know, artificial intelligence. I feel like I was talking with a real person. But yet I didn't feel that uncomfortable like, you know, I would talking with someone, one on one."

The focus on heavy-drinking adults recruited through online social networks proved particularly effective, with 85.1% of participants finding the agent useful, e.g. because "he explained things that I did not know" and "it helped me recognize that I have a problem." This demonstrates that VHAs can effectively serve screening and brief intervention functions, potentially addressing the significant treatment gaps in substance abuse care [20]. The comment that "The questions she asked me, no one had asked me before and helped recognize my drinking problem" highlights how VHAs can reach individuals who might not otherwise receive evidence-based interventions.

Table 10. User's perception/encoding of, and user's socio-emotional responses to, the virtual agent counselors' **appearance social cues** (realism, clothing/style) to support the metaphor of a human adult empathic counselor: **VHA's appearance social cues** shown in Column 1, are perceived and encoded by users as **VHAs' appearance social signals** shown in Column 2, which in turn triggers **user's socio-emotional responses** shown in Column 3, with **excerpt examples from interviews** shown in Column 4, and weighted averages (M) (cf. Figure 7).

User's perception/encoding of, and user's socio-emotional responses to, the VHAs' APPEARANCE social cues			
VHA's appearance social cues (see Section 5, & Figures 2, 5, 6)	Users' perception and encoding of VHA's appearance social cues as VHA's appearance social signals	User's socio-emotional responses to VHA's appearance social signals	Sample Excerpts of users' socio-emotional responses to VHA's appearance social signals (M=4.3)
Realism: - 3-Dimensional, - <i>HumanBasic</i> [110] level of realism, - human-like	looks like a regular person, as realistic as it can be & still cartoony, computer generated, anime-like character, not like video games	appropriate for application, not creepy, really liked, friendly, like father figure, easy to talk to about private matters	- <i>She was friendly, she didn't creep me out or anything.</i> - <i>Well, I really liked him. My dad, he's still living but he's not like I said he is kinda religious. so I don't talk to him about things like this. So that's why I went with Luke because he kind of even looked like my dad [...]. my dad is just a really wise guy a very highly intellectual.</i>
Attractiveness (for Western[ized] populations): face : symmetry, averageness;	biological attractiveness	helpful	- <i>Um, I think it definitely helps to have someone that you know, is not, I don't want to say it's like good looking, but just not *inaudible* like not bad looking.</i> - <i>The person, I thought they were pretty attractive looking.</i>
Clothing and business casual style (for Western[ized] populations): - female-looking: dress, blouse & slacks, skirt & sweater; Western hair styles; - male-looking: slacks & shirt & tie or not & casual jacket or not - Western hair styles, Rastafari braids, bald with glasses	braids, Rastafari, well dressed, casual style	attached to, influential, professional, relatable, laid back, intelligent	- <i>One, he, I was attached to him because he had braids and he looked, he seemed to be more influential.</i> - <i>I liked how he was well dressed. It seemed like his coat wasn't buttoned if I remember right. So he was well dressed and professional, but you know just a little bit relatable like laid back casual like, Hey, you can talk to me.</i>

9.2 Therapeutic Relationship and Social Presence Design (RQ1)

The quantitative findings demonstrate that virtual agents can achieve meaningful therapeutic relationships, with 68.9% of participants reporting that it sometimes felt like talking to a real person, and 90% perceiving the agent as warm. These metrics are also supported by qualitative descriptions that highlight the mechanisms underlying therapeutic engagement.

9.2.1 Personal Qualities: Based on the collection of 77 individual social cues we implemented (added up in the last eEVA-4AUD column in Tables 15 and 16), and on the multimodal of communication we created from combining verbal, kinesics, and proxemics social cue, adult alcohol users' socio-emotional responses were similar to those of clients with successful counselors (e.g., "very caring," "very calm," "reassuring," "very attentive," "non-judgmental"). The Godspeed questionnaire results validate the agent's therapeutic qualities, with participants rating the agent highly on warmth 90%, responsiveness (89%), knowledgeability (92%), and being nonjudgmental (84%).

9.2.2 Appearance and Realism: The 3D human-like VHAs rendered with realism level à la *HumanBasic* [110] looked like real persons to adults receiving BMI, were found appropriate, elicited positive feelings, and avoided falling into the uncanny valley [123]. Most participants found the VHAs looked realistic, "like a regular person," and "appropriate for the application." Our findings corroborate previous research showing that realistic, yet not photorealistic rendering, is more appropriate than cartoon-like stylized characters for healthcare applications [149, 152].




Table 11. User's perception/encoding of, and user's socio-emotional responses to, the **virtual counselors' nonverbal (NVB) social cues** that support the metaphor of a human empathic counselor: **VHA's NVB social cues** shown in Column 1, lead to users' interpretations of these social cues as **VHA's NVB social signals** shown in Column 2, and to **user's socio-emotional responses to VHA's NVBs** shown in Column 3, with **excerpt examples** shown in Column 4, with weighted averages (M) of coder's ratings (1 very negative comment - 5 very positive comment) (cf. Figure 7).

User's perception/encoding of, and socio-emotional responses to, the VHAs' NON-VERBAL (NVB) social cues			
VHA's NON-VERBAL (NVB) social cue design features (see Section 5 and Figures 4 and 6)	Users' perception/encoding of VHA's NVB social cues as VHA's NVB social signals	Users' socio-emotional responses to VHA's NVB social signals	Excerpts examples of users' socio-emotional responses to VHA's NVB social signals (M=3.69)
In all states: speaks with lip synchronization , blinks (15-20 / min.), breathes	-	-	-
Greeting: types while seating on computer on its desk, gazes toward the user "standing" at the door, walks to change proximity to user from social sphere to personal sphere closer to user, sits down at eye-level with user (different camera angle)	ability to work, smartness, office features, user's first person view of office room while standing	really liked NVBs a lot, really cool, VHA is smart, interaction is similar to coming to an appointment	- <i>I really liked that a lot. i don't know why. I just thought it was so cool like here he is he's working on his stuff. He is a smart dude, he's got a nice office. And then I come in, it's like coming into an appointment. So I thought that was really cool.</i>
Neutral (ongoingly): while seated , with open body language with uncrossed arms and uncrossed legs (open legs for males, legs together for females), hands on thighs , VHA speaks to user with head gestures and eyebrow movements , gazes toward the user, who has "seated" in front of VHA	realistic, open body language	really really impressed, interaction is like therapy, VHA is very open, non confrontational, and non judgmental	- <i>Yeah, I thought it was good. You know, reminded me of like, what a therapist to do movies and stuff, and very non confrontational. I thought it was good.</i> - <i>I was really really impressed with the body language design. It was like very open body language.</i>
Asking: questions user with single hand formless flick ; Listening: waits for the user to respond to the question, leans forward on its seat, maintains a leaning posture until the user answers, with one of the three different expressive nods (neutral, concerned, smiling) (depending on the user's answers, returns to an upright sitting position; Explaining: speaks with two-handed flip gesture & two-handed contrast gesture	active listening subtle gestures, coordinated verbal and NVB active listening behaviors	liked everything, VHA is really attentive, and listening to user, VHA is validating user, user feels good about the VHA	- <i>Um, I pretty much liked everything. I like her little gestures. It's seemed like she was really attentive she gave like little different gestures to let me know kind of that she was listening. so, I liked that.</i>
Greeting upon user's return from a previous session re-greets user (see Greeting above), while verbally reassures user that it remembers between sessions all user's answers, and where they left of in the previous session	ability to greet user, ability to remember user	really liked VHA, similar to returning to an appointment, VHA is reassuring	

9.2.3 *Clothing and Professional Appearance:* We dressed the virtual counselor in business casual attire (e.g. dress shirt, no tie) to balance professional credibility with approachability, consistent with the collaborative counseling stance of MI, rather than an authoritative physician role. Users confirmed this achieved the intended effect, describing the counselor as "professional," "intelligent," and "relatable," noting the attire conveyed "well dressed and professional, but... laid back casual like, Hey, you can talk to me." This contrasts with another study's recommendation of white coat and stethoscope for virtual counselors [138], which emphasized hierarchical authority rather than the partnership model central to MI.

9.2.4 *Active Listening Behaviors:* The integration of active listening behaviors was successful, with participants noting that the agent was "really attentive" and "listening," and that the "little different gestures" made them feel "validated." The VHA's nonverbal behaviors (M = 3.69) were perceived as signaling openness and triggered

Table 12. User's perception/encoding of, and socio-emotional responses to, the VHA's **diversity social cues** to support the diverse counselors metaphor.

User's perception and encoding of, and socio-emotional responses to, the VHAs' DIVERSITY social cues			
VHA diversity social cue design features (see Section 5.1.1 and Figure 2)	Users' perception and encoding/interpretation of VHA's diversity social cues as VHA's diversity social signals	Users' socio-emotional responses to VHAs' diversity	Sample Excerpts of Users' socio-emotional responses to VHA's social cues
Race/Ethnicity (M=4.46) 	race of VHA	liked, definitely a plus, more comfortable with VHA of similar race as self	<ul style="list-style-type: none"> - <i>I thought it was cool. Um, I obviously I chose somebody of my own race [...]</i> So Um, I think I think it's definitely a plus... Not like not only white people have drinking problems. - <i>So, um, I've been to therapists too. And I think when I choose someone, I kind of want someone that's similar gender and ethnicity for me. So that is why I chose someone that was like, also Asian and I think this one was Japanese.</i>
Age (M=4.55) 	age of VHA	liked VHA's closer to self in age, older VHA as wise, good therapist, father figure	<ul style="list-style-type: none"> - <i>Maybe I picked her because she's like closer to me, by far as maybe ethnicity and maybe age.</i> - <i>I just picked the older gentleman because it seemed like he would be more like a good therapist.</i> - <i>I probably kind of have a fatherly figure, man, I see my dad as nice, but I think maybe a lot of people do associate age with wisdom.</i>
Gender (M=4.41) 	gender of VHA	quite important, more comfortable with female and minority, more in common, interesting and surprising to have choice	<ul style="list-style-type: none"> - <i>[Having diverse choices] is quite important. I really emphasize I'm choosing someone that's also female because I feel more comfortable and also someone that is a minority.</i> - <i>I figured that we had much more in common than other agents.(...) Appearance, age.... I would say gender too.</i> - <i>I thought it was interesting that there were a lot of agents to choose from. For something like this, I didn't expect to choose from many agents.</i> - <i>Um, I felt more comfortable talking to him than I did with the rest of them.(...) I didn't even know his name, I just was like "this is the one I want to talk to."</i>

responses that the VHA was "non judgmental" and "non confrontational." One user said: "I was really really impressed with the body language design. It was like very open body language."

9.3 User Acceptance and Diversity Considerations (RQ3)

9.3.1 Self-disclosure: Quantitative findings show that 90.5% of participants were comfortable disclosing drinking information to the digital health assistant, with 83% reporting greater comfort compared to disclosing to their medical doctor. This preference for virtual disclosure is explained through qualitative insights: "The assistant gives me a level of trust that I don't have with other humans," "It is more easy talking to (the health agent) than to a real person," and "I don't feel like they are judging me." The high level of disclosure comfort represents an advantage for substance abuse treatment accessibility, which confirms prior research on self-disclosure with agents [59].

9.3.2 Diversity and Choice: . Adults with AUD had unanimous positive socio-emotional responses to having the option to choose diverse VHAs in terms of race, gender, and age. They considered this both welcome and important, with one participant stating: "I like that you can choose which, you know, which one you wanted, younger, older, man, woman, different races." Results indicate that 53% of participants selected VHAs they perceived to share their own race, corroborating homophilic responses [113]. Participants often chose agents similar to themselves: "I kind of want someone that's similar gender and ethnicity for me. So that is why I chose someone that was like, also Asian." However, some users chose the "older male" character for perceived wisdom

and experience, suggesting homophilic responses aren't always applied regarding age. One participant explained: "I just picked the older gentleman because it seemed like he would be more like a good therapist." Our findings contrast with recommendations that virtual agents in healthcare could be designed as androgynous to avoid stereotypes [158, 160], pointing to a need for more research.

9.3.3 Voice and Speech Recognition: While participants found the agent physically appealing (74%) with a pleasant voice (82%), voice speed received the lowest rating ($M = 2.19$), with most users finding it "too slow." Additionally, only 67.6% felt the agent understood them when speaking, primarily due to microphone configuration issues rather than technology limitations. Findings suggest that verbal reflective listening social cues are needed to avoid repetitiveness of simple acknowledgements, and voice speed adjustment options should be provided.

9.4 Technical Implementation and Future Enhancements

The quantitative evaluation revealed both strengths and challenges in technical implementation. While 86.5% of participants found the agent easy to interact with and learned to use quickly, technical challenges emerged in specific areas.

9.4.1 Voice Recognition and Synthesis: Although most users found voices "pleasant" and "natural," some found them "too text-to-speechy" and "robotic." The monotonous voices without intonation variations contributed to perception of low emotional expression. Research shows that facial expressions are predominant over vocal expressions [?], which may explain why overall user experience remained high despite voice issues. Future research is needed on voice qualities.

9.4.2 Sensing Social Intelligence: While alcohol users fully accepted VHAs that could encode (express) subtle social cues, not all users are ready for VHAs' ability to decode (recognize) their own social cues. Mixed responses to potential facial recognition ($M = 3.43$)⁴ highlight privacy considerations, with some expressing concerns: "I don't really feel like comfortable with like facial recognition stuff in general." These mixed reactions suggest that research studies are needed to identify users' needs and preferences in terms privacy concerns and acceptance levels of socially sensing virtual agents, as also suggested in a review of desired features for VHAs that questioned whether sensing VHAs are necessary to for users to experience VHA's empathy [165].

Despite some technical challenges, 74.3% of participants felt they could have a meaningful conversation with the agent, and 85% expressed willingness to interact with it again. The high engagement scores (mean = 2.86, SD = 0.96) and finding that 90% believed the system could be adapted to their needs suggest strong potential for sustained use continuation.

9.5 Therapeutic Environment Design (RQ2)

The creation of the therapeutic "soft room" environment successfully supported user engagement, with participants unanimously rating the virtual office environment positively ($M = 5.0$ on a 5-point Likert scale). Qualitative responses validate our research-based environmental design principles, with users describing the office as "really comforting," "very comfortable," "very therapeutic," and "not bland nor overdone." One participant noted: "Like I felt like the room was a very good, kind of a therapeutic setting type room."

The environmental design successfully conveyed professional therapeutic signals while maintaining approachability. Users perceived environmental cues as signaling both an "office environment" and a working space where the VHA "is working on his stuff." The "clean setting" was particularly appreciated, with one participant explaining: "I thought they were uh in a clean setting. The setting was cool cuz I really liked, you know, to be in a nice setting myself."

⁴Reminder: M represents the weighted averages of weights assigned by coders on a 5-point Likert scale based on the valence of user's interview comments (1 for very negative comment, 5 for very positive comment).

Table 13. User’s perception/encoding of, and socio-emotional responses to, the VHA’s **auditory social cues** that support the metaphor of a human empathic counselor: **VHA’s vocal social cue design features** shown in Column 1, lead to users’ encoding/interpretation of these social cues as **VHA’s vocal social signals** shown in Column 2, and to **user’s socio-emotional responses to VHA’s vocal signals** shown in Column 3, with **excerpt examples from interviews** shown in Column 4, and weighted averages (M) (cf. Figure 7).

User’s perception and encoding of, and socio-emotional responses to, the VHAs’ AUDITORY social cues			
VHA’s vocal social cue design features	Users’ encoding/interpretation of VHA’s auditory social cues as VHA’s vocal social signals	Users’ socio-emotional responses to VHA’s vocal signals	Excerpts examples of users’ socio-emotional responses to VHA’s verbal tone (M=2.9), and voice speed (M=2.19) social signals
Voice quality : 4 voices from <i>ReadSpeaker</i> : two female voices (Julie, Ashley) and two male voices (Paul, James); voices were respectively matched with female- and male-looking characters.	text to speech, no intonation, gender of voice, biological ethnicity of voice	robotic, monotone, too artificial, too text to speechy, has no emotions, pleasant, pretty realistic, natural, matches VHA’s appearance, attribution of gender and ethnicity (urban) of voice	<ul style="list-style-type: none"> – <i>Nothing [to change] really, other than, you know, the robotic voice.</i> – <i>They had no emotions. You know? [...] the tone of the voice was very I guess animated or monotone.</i> – <i>I thought [the voice] was natural to me. Like being there with that person.</i> – <i>His voice sound urban. like he is- From my, from he’s probably black just like me. So that’s probably one of the reasons why I’m picking, he sounded like me.</i>
Verbal tone : simple verbal reflections, added at times after user’s answer (see Section 5.4 on page 28)	verbal reflection repetitions	a little annoying after a while, validating what I was saying, good	<ul style="list-style-type: none"> – <i>The voice was fine too. Like the only real problem I had was the repetition of “oh, okay.”</i> – <i>So yeah, ... he would be like, “mm hmmm” like, “thanks for sharing that.” I don’t remember what he would say. so yeah body language, that kind of validation of what I was saying, and [it] feels good yeah.</i>
Voice speed	constant voice speed, slow speed	not as fast as real human conversations, way too slow, perfect because clear and slow	<ul style="list-style-type: none"> – <i>Slow in as like when the virtual person was talking it didn’t seem to be as fast as a real human conversation would be.</i> – <i>I think it was perfect because uh,. Even without the closed captioning, I mean, he spoke clearly, slowly but not like speaking slow, but we completely understand everything that was being said or asked that was cool.</i>
Suggested improvements : emotional prosody, faster/adjustable speed, less repetitions	–	–	<ul style="list-style-type: none"> – <i>[Going back options] are also very useful. So I guess like, playback speed, like a “faster” playback speed? Yeah. Yeah. Yeah.</i>

Our findings differ from previous research [138] that found medical examination rooms professional for counseling. Different metaphors (medical examination vs. counselor’s “soft room”) methodologies and user groups led to different findings regarding appropriate design features. Our findings demonstrate that virtual therapeutic environments can successfully incorporate established design principles from environmental clinical psychology research [171, 193].

10 Extended Taxonomy

Lastly, to facilitate systematic accumulation of knowledge and comparison of various virtual agents, as suggested by its authors, we propose an extension to Feine’s et al. [51] taxonomy, by adding levels of abstraction, individual social cues, and a column identifying the inventory of individual social cues in our eEVA-4AUD implementation. We formatted our extensions into the tabular format shown in Tables 15 and 16, in which (1) we created a column (labeled in teal font) for levels of abstraction or values that social cues can have, (2) we added social cues to the original taxonomy (also color-coded in teal font), and (3) we added a column detailing the social cues used implemented in our current eEVA-4AUD virtual health agent. Given that we added social cues within the same

Table 14. User’s perception/encoding of, and socio-emotional responses to, the VHA’s **office environment** that support the metaphor of a human empathic counselor’s office: **VHA’s office social cue design features** shown in Column 1, lead to users’ perceiving and encoding these social cues as **VHA’s office social signals** shown in Column 2, and to **user’s socio-emotional responses to the VHA’s office** shown in Column 3, with **excerpt examples from interviews** shown in Column 4, and weighted averages (M) (cf. Figure 7).

User’s perception/encoding of, and user’s socio-emotional responses to, the VHAs’ OFFICE ENVIRONMENT social cues			
VHA’s virtual office social cue features (see Section 5.1.3 & Figures 4 and 6)	Users’ perceived and encoded of VHA’s office social cues as VHA’s office social signals	User’s socio-emotional responses to the virtual office	Sample excerpts of user’s socio-emotional responses to the virtual office social cues; user experience with the office M=5, office realism M=5, office views M=4
Aesthetically pleasing decor: curtains, decorative rug, poster of nature, plants, slightly hued walls, flower vase, mirror, color-coordinated furniture	objects in office	really comforting environment, good for one one one conversation, socially attractive room, very good therapeutic room, very comfortable, not blend nor overdone	<ul style="list-style-type: none"> – <i>I think it was good. Like it was really comforting.</i> – <i>It was cool. It was like sitting in an office environment or whatever and actually having a one on one conversation.</i> – <i>Like I felt like the room was a very good kind of a therapeutic setting type room.</i> – <i>Very comfortable. It wasn’t, you know, too overdone or too bland and it was just like any other office, you would walk into.</i>
Order and neatness	objects in office in expected location	clean, nice setting, real office	<ul style="list-style-type: none"> – <i>I thought they were uh in a clean setting. The setting was cool cuz I really liked, you know, to be in a nice setting myself. (call volume drops).</i>
Non formal larger work space with two part: (1) desk, computer, office chair, table lamp, ceiling fan, diplomas behind desk, books in bookcases; (2) ample seating for user: 3 padded chairs and 1 loveseat, coffee table, vases, tissue box	work space, office	ability to work, smart, nice office	<ul style="list-style-type: none"> – <i>I just thought it was so cool like here he is he’s working on his stuff. He is a smart dude, he’s got a nice office.</i>
Suggested improvements: warmer light, different inside views of room, more outside views, take a walk outside with the VHA while he talks	–	–	<ul style="list-style-type: none"> – <i>I feel like it’d be- it could be more like a more relaxing atmosphere like warm lighting and stuff. [...] Usually when I go to see my therapist they have a lamp on that has a warm glow to it and stuff like that.</i> – <i>So being able to relate different angles of the room might have even made the bot seem like more like a real person [...]</i>

hierarchical categorical structure of the original taxonomy, we did not change its structure and its validity is maintained. We explain below the rationale for our taxonomy extensions.

10.1 Adding Levels of Abstraction for Social Cue Characteristics

To enable agent designers to document how they implemented their individual social cues, we broke down the social cues into different potential fine-grained design characteristics, based on our knowledge of existing virtual agents. Table 15 shows the taxonomy of visual social cues and a *Value/Characteristics* column describing possible values or characteristics each social cues can be designed to have.

10.1.1 Visual Appearance. In the **Visual Appearance** sub-category of the original taxonomy, the [Photo-]realism social cue is described as "the appearance of the agent ranges from extremely photorealistic to a comic appearance". While the original taxonomy used the term Photorealism to refer to the appearance level of realism, we suggest to rename that social cue as **Realism**, and to add a scale for realism levels, which can include a photoreal appearance.

Using fine grained scales, borrowed from computer graphics research on virtual characters, can help better describe and document the degree of realism of a VHA. Given the large number of existing rendering styles, we acknowledge that it is difficult to assign a realism scale to a virtual character. However, we posit that to facilitate consistency across the research (e.g. regarding claims about evaluation of a character based on its level realism),

Table 15. Extended Taxonomy of **Visual** social cues (expanded from [51]), additions shown in teal: Value/Characteristics column describing single cue characteristics; eEVA-4AUD column with check marks for cues implemented in our design.

Visual social cue		Value/Characteristic	eEVA-4AUD
Appearance	2D/3D	2D	✓
		3D	
	Realism	ToonPencil (comic-like)	✓
		ToonFlat	
		ToonShaded	
		ToonBare	
		ToonCG	
		Human III	
	Humanlikeness	Human Basic	✓
		Human SSS	
		Human HQ1	
		Human HQ2	
	Viewing Angle	Low	✓
		Neither	
		High	
	Facial Feature	Frontal (0°, horiz)	✓
		Slight angle (15°, horiz)	
		Moderate (30°-45°, horiz)	
		Eye level (vert)	
		Neutral Expression	
	Gender	Slight smile (30-50% intensity)	✓
		Warm Smile (60-80% intensity)	
	Age	Female	✓
		Male	
	Color/Race/Ethn	Ambiguous	✓
		[26-39] Adult	
		[40-64] Middle-aged	
	Attractiveness	>64 Older Adult	✓
African American			
Clothing	Asian	✓	
	Caucasian/White		
	Native American/Alaska		
	Nat. Hawaiian/Pacific Island.		
Name Tag	Hispanic or Latino	✓	
	Yes		
Facial Feature	RGB color	✓	
	Low		
Attractiveness	Neither	✓	
	High		
Clothing	Casual	✓	
	Business casual		
	Business formal		
Name Tag	White coat	✓	
	Stethoscope		
Facial Feature	Yes	✓	
	No		

Visual social cue	Value/Characteristic	eEVA-4AUD	
Kinesics	Facial expression	Smile (AU6+12)	✓
		Concern (AU4+2+4+5+7)	✓
		Subtle surprise (AU1+2+5)	
		Interest (AU1+2+direct gaze)	
	Viseme movement	Other AU combinations	
		Yes	✓
	Eye brow mvt.	Raise	✓
		Furrow	
		Flash	
	Eye movement	Unilateral raise	✓
		Direct gaze	
		Gaze aversion	
		Gaze at object	
	Head movement	Idle eye movements	✓
		Blink AU 45 (15-20/min)	
Nod			
Positive nod (nod+smile)			
Negative nod (nod+concern)			
Arm and hand gestures	Double nod	✓	
	Shake		
	Tilt		
	Turn		
Posture shift	Arms uncrossed	✓	
	Typing		
	Formless flick		
	2-handed flip		
	2-handed contrast		
	Twist		
Leg movement	Breathe	✓	
	Lean Forward		
	Lean Back		
Proxemics	Environment	Legs uncrossed (sitting)	✓
		Walk	✓
		Sit-to-stand	✓
		Stand-to-sit	✓
	Conversation distance	2D	✓
		3D	
		Spacious office room	
		Desk	
		Computer	
		Chair	
		Sofa	
		Window	
		Outdoor nature views	
		Clock	
		Tissue box	
Plants			
Diplomas			
Diffuse lighting			
Art			
Emoticons	Examination room	✓	
	Medical equipment		
	Bed		
Typefaces	Bench	✓	
	Intimate		
	Personal		
Traditional UI elements	Social	✓	
	Yes		
	Yes		
Visualizations	Yes	✓	
	No		
No. of implemented visual social cues in eEVA AUD		58	

using realism scales from computer graphics literature can help get closer to some research consistency within the field.

We chose to use a scale of *10 different rendering styles*, ranging from comic-like to photorealistic proposed by McDonnell, et. al. [110] shown in Table 15. McDonnell et al. [75] more recently described 8 additional ranges for degrees of realism with yet finer grained levels of rendering that can be achieved with state of the art graphics techniques, reaching higher photorealism. Given that such photoreal graphical models have not yet been used to build virtual health agents from our understanding of the current literature (and research is needed to know whether that would be desirable), we therefore expanded the taxonomy with the 10 possible values discussed in [110], which is sufficient to describe current 3D virtual health agents. As more realistic VHAs are designed and studied, researchers can extend/replace these scales with realism values that best describe the level of realism of their 3D characters.

We added recommended values for the virtual agent's **facial feature**, defined as the static and non-changing characteristics of the agent's face: neutral expression, slight smile (30-50% intensity, warm smile (60-80% intensity), that can be used to first introduce the agent. In the counseling domain, we want to portray positive emotions, so we used the facial feature to first introduce the agents.

We renamed the original Color social cue to enable designers to document the human demographic groups under the **Color/Race/Ethnicity** social cue, and included fine-grained value for these. A recent review of 72 virtual agents reported that the ethnicity of 67% agents were white-skinned, 4% tan-skinned, 6% dark-skinned, 4% Asian, and 2% Asian [102], demonstrating designers' tendency to focus on white virtual agents. Given that skin color as a physical characteristics is more related to race than to ethnicity, we used *Color/Race/Ethnicity* to enable designers to document human-like agents from different demographic racial or ethnic groups they might want agents to portray. Because our study was conducted in the USA, we used the racial and ethnic categories from the US Census Bureau [185] (which we also used to collect our participants' demographics) to provide potential values for that cue. These are not exhaustive, and additional values can be added to represent other groups as needed. We also added a value for colored agents (e.g. pink, green) that can be documented with RGB values, although again this would not be feasible for metaphorical design of an adult counselor metaphor.

Similarly the recent virtual agent survey reported that 74% were adult, with only 2% were older adults, so we added values to represent most **age** groups. We only included adult ages because our work is focussed on the adult counselor metaphor. For studies with teens, which some study indicates they prefer same age virtual peer coaches that look like them (in addition to a virtual adult counselor in a different role) [94], younger ages can be added.

We added characteristics for **Clothing** that have been used by the very few studies on clothing, including: business casual (as in [98] and in this study), business formal, or white coat with stethoscope which can be useful to design agents based on the metaphor of counseling in primary care settings [100, 138]. For studies on same age virtual peer coaches [94], casual clothing might be preferred.

10.1.2 Visual Kinesics. For the **Visual Kinesics** social cue category **Facial Expression** social cue, we added the names of FACS-based expressions that our evidence-based approach and current findings point to as most useful in the counseling domain, depending upon the specific context of the dialog: smile, and concern which we used for this study; subtle surprise, and interest might also occur depending upon the dialog. Other expressions derived from combinations of AUs might also be useful, again depending upon context.

We added **Eye brow movements** values that might be relevant in the counseling context: raise (or "eyebrow raise") that might signal surprise, interest, questioning, or recognition; furrow (or "eyebrow furrow") that might signal confusion, concentration, concern, or requesting clarification; flash (or "eyebrow flash"), a brief raise-and-lower that might signal recognition, greeting, or communicative intent; and unilateral raise, one eyebrow raised that might signal skepticism or questioning.

We added **Eye movements** values that are most likely in dialog: blinks, direct gaze because eye contact is associated with enhanced empathy and alliance in counseling, as well as gaze aversion to reduce emotional

intensity, or for turn-giving; gaze at object, e.g. for reviewing information; and idle eye movements for avoiding frozen stare.

As discussed in Section 5.1.2, we added values for: **Arm and hand gestures** that can be characteristics of full-bodied animated virtual agents inhabiting an office environment such as arms uncrossed (communicating openness), typing (as a working professional), and discussing topics with users with the help of formless flick, 2-handed flip or 2-handed contrast gestures, and resting (as our current agent does); for **Head movement**: nod, positive nod, negative nod, double nod, shake, tilt, and turn; and for **Posture shift**: twist, breathe, lean forward, lean back. These are all important social cues for active listening.

10.1.3 Visual Proxemics. As discussed in Section 5.1.3, in the **Visual Proxemics** sub-category, we: renamed the Background social cue to **Environment** to capture the fact that 3D virtual agents can be portrayed in front of a 2D background image [98], but that most often they evolve in a 3D environment (as with our current virtual agent); we specified a set of objects that have been used to model a counselor’s office (as our current agent), or those found in an examination room with standard medical equipment (a bench, a bed, and medical equipment) for counseling performed in medical offices [138]; and we provided values for **Distances** inspired by proxemics theories: intimate, personal, and social.

10.2 Adding Individual Social Cues

Since VHAs used in eEVA-4AUD were implemented with social cues that were not included in the original taxonomy, we added individual social cues to some of the original categories: in the Visual category, we added to the Appearance, Kinesics, Proxemics, and Computer-Mediated Communication (CMC) sub-categories; and in the Verbal category, we added to the Content sub-category.

In the **Visual Appearance** sub-category, we added the **Viewing Angle** social cue, defined as *the horizontal and vertical angle from which the virtual counselor’s face is presented*, e.g. horizontal slight angle (15°), full frontal (0°), or vertical angle at eye level. These can be manipulated with camera angles to produce different appearance effects.

In the **Visual Kinesics** sub-category, we added two social cues: **Viseme**: we portray realistic lip synchronization by mapping visemes to phonemes; and **Leg movement**: as discussed in Section 5.1.3, we added leg movements with the following values: legs uncrossed, walk, sit-to-stand, stand-to-sit.

In the **Visual CMC** (computer-mediated communication) sub-category, we added cues for **visualizations** that agents such as ours use to present data to users (in this case about their drinking), and **UI elements** for traditional UI elements that are needed to complement and augment VHA’s communication, e.g. microphone on/off icon.

Table 16. Extended Taxonomy of **Verbal, Auditory, and Invisible** Single Social Cues (expanded from taxonomy by [51]), with additions shown in teal: new cues; new eEVA-4AUD column with check marks for cues implemented in our design.

Social cue	Value/Characteristic	eEVA-4AUD
Verbal	Acknowledge	✓
	Ask to start/pursue dialog	✓
	Ask	✓
	Excuse	✓
	Explain	✓
	Feedback	✓
	Greetings and farewells	✓
	Joke	✓
	Opinion conformity	✓
	Praise	✓
	Refer to emotion	✓
	Refer to past	✓
	Reflection	✓
	Self-disclosure	✓
	Self-focused questions	✓
	Small talk	✓
	Thanking	✓
	Tips and advice	✓
	Abbreviations	✓
	Formality compliant	✓
Style	✓	
Lexical diversity	✓	
Sentence complexity	✓	
Strength of language	✓	
Auditory	Voice qualities	✓
	Gender of voice	✓
	Pitch range	✓
	Voice tempo	✓
	Volume	✓
	Grunts and moans	✓
Vocalizations	Laughing	✓
	Vocal segregates	✓
	Yawn	✓
	First turn	✓
Invisible	Chronemics	✓
	Response time	✓
	Haptics	✓
	Tactile touch	✓
	Temperature	✓
No. of implemented eEVA verbal, auditory, invisible social cues		19

As discussed in Section 5.4, in the **Verbal Content** sub-category, we added the **Acknowledge** social cue, e.g. "Okay.", "I see.", that is often used in conversations, and in particular in counseling sessions, to build therapeutic alliance combined with nonverbal active listening behaviors. We also added the **Ask, Explain, Feedback** social cues that are underlying verbal actions afforded by BMIs, the **Refer to emotion** to document when an agent refers to the user's perceived or inferred emotions from user's verbal or nonverbal messages. This characteristics can enable an agent to verbalize empathic messages, e.g. when providing feedback empathically about individuals' health conditions or lifestyle. Finally, we added the **Reflection** social cue that is one technique to express empathy by reflecting on what the user shared, or in its simplest form by rephrasing or repeating what the user said to confirm that it has been considered and understood [198]. Motivational interviewing, however, uses a wide variety of therapeutic complex reflections [121] that future VHAs should provide.

10.3 Adding Comparison Checks

Lastly, we added the *eEVA-4AUD* column to indicate with check marks which individual social cues with their values we implemented in the current implementation of our our virtual health counselor eEVA-4AUD. This knowledge can facilitate consistent comparisons of our agent with others, as was exemplified with three different agents earlier in [51]. Additional columns can be added to compare which social cues and which values were implemented in different agents.

11 Limitations and Future Directions

11.1 Limitations

While these findings represent significant progress, several limitations must be acknowledged.

Firstly, despite mixed ratings on visual realism of the virtual agents, participants found the agent physically appealing (71.6% agreement), interesting (88% agreement), and responsive (89% agreement), and were willing to use the system again (85% agreement). These findings align with the I-PEFiC model (see Section 5.1.1), which predicts that realism contributes minimally to engagement compared to facial familiarity and aesthetics, and does not significantly influence use intention. Moreover, photorealistic agents may be inadvisable for counseling contexts: the uncanny valley effect shows that near-human photo-realism can trigger discomfort, reduce trust, and create confusion, whereas realistic agents avoid this risk while maintaining effectiveness. Further research is needed to identify desirable levels of realism for the counseling domain, which might depend on user population. It should be acknowledged that while an initial novelty effect could play a role in the social presence felt [141], the aspects of the system described represent broader areas of design consideration, with potential examples from eEVA-4AUD to demonstrate how such features can be integrated.

There was also variability around the agents' interpretations and preferences. The agents were initially designed and assessed with primarily Western (and specifically North American) standards of appearance and work roles. All conversations with the agents were held in English. Such findings may not be globally extensible, and preferences, and familiarity can impact rapport formation. Although one participant mentioned that "*No, I think they covered pretty much all age groups and genders and nationalities*", we cannot make that claim: human demographic uniqueness is based on infinite combinations. The cultural and linguistic background of participants was limited, and we cannot say their preferences reflect all users, nor that all types of potential looks were represented in our agents. Whereas this work represents a first step to broadening the options available as users engage with VHAs, research is needed to further discover users' agent's preferences, including about androgyny.

Although the evidence-based behavior change intervention on alcohol use disorder that the VHA delivered is lengthy and could be completed in multiple sittings (which we limited to one month), a longitudinal study would enable us to provide guidance with generalizable results on long-term interactions. The study sample was limited

to heavy-drinking adults recruited through online social networks, which may not generalize to other clinical populations or delivery contexts.

Technical improvements needed include: enhanced speech synthesis with emotional intonation, smoother animations for nonverbal behaviors, and more robust voice recognition systems. The current model can be enriched with more facial expressions [131, 140] and expanded nonverbal repertoires while maintaining the successful active listening framework.

Our extension of the taxonomy might not encompass all additions needed to portray virtual agent as counselor extenders, but represents a continuation of the systematic discussion of individual single social cues for detailed agent design started by the author of the original taxonomy [51].

11.2 Clinical Practice and Implementation Implications

The integration of quantitative acceptability data with qualitative user feedback demonstrates that virtual counseling systems can achieve both high user satisfaction and therapeutic fidelity. The finding that 85% of participants would interact with the agent again, combined with 90% believing the system could be adaptive to their needs, suggests good potential for sustained engagement and treatment continuation.

The evidence supports the conclusion that virtual agents designed as counselor extenders have the potential to improve access to evidence-based health interventions within the healthcare ecosystem, e.g. initial stepped care, prevention intervention, booster sessions, just-in-time interventions. Human clinicians should not be replaced, these systems should extend therapeutic reach to underserved populations, provide consistent evidence-based interventions at scale, and potentially serve as a first point of contact for individuals who might not otherwise seek treatment due to stigma, accessibility, or resource constraints. The convergence of workforce shortages and demographic mismatches underscores the potential for virtual counselors to provide culturally representative behavioral and mental health support to underserved communities.

The systematic approach to design and evaluation, validated through both quantitative metrics and qualitative user experience, offers a replicable framework for future virtual health counseling interventions [96]. Albeit our system is a non-commercial research prototype, the findings that participants viewed the agent as providing "reliable information and statistics" while being "practical, easy to use, and guides the person on what to do without forcing us to make a final decision" suggest a successful prototypical implementation of evidence-based, BMI therapeutic approaches.

12 Concluding Overarching Findings

Our work demonstrates that it is feasible to implement and deliver complete, evidence-based Brief Motivational Interviewing interventions using virtual agents as counselor extenders. The qualitative evidence shows that participants recognized the client-centered approach as a usable and useful intervention. The finding that 93% were comfortable disclosing to the virtual agent (with 83% preferring it over physicians) addresses a critical implementation challenge. This suggests virtual agents could serve as accessible first points of contact for populations who avoid traditional care due to stigma. Our 86% enjoyment rate, 85% willingness to interact again, and high ratings on therapeutic qualities (90% warmth, 92% knowledgeability) suggest virtual agent as counselor extenders - even without privacy-concerning automatic emotion recognition to simulate adaptive nonverbal mirroring and richer empathic communication - can achieve meaningful therapeutic relationships, challenging assumptions that technology-mediated care is inherently impersonal.

The potential use of adaptive social intelligence, in which a virtual agent has access to and interprets users' emotions from facial expressions, and the privacy concerns it raises needs to be investigated to assess users' acceptance of such virtual agent technologies in the context of counseling. Identifying the best level of agents' realism for best engagement outcomes and satisfaction outcome may vary based on a variety of factors, including

age, gender, computer game experience, among others, and needs to be further investigated. Greater variety in virtual health agents as counselors can offer significant value and homophily from those engaging. This variety could aid in rapport formation and better virtual health agent adoption and resulting health outcomes, and should be explored further in future studies.

The use of a taxonomy of social cue to identify and document each individual feature, of evidence-based ‘soft room’ therapeutic environment office designs, and of user-centered evaluation using mixed methods created a comprehensive system rather than fragmented components. This holistic approach might be useful for future designers of virtual health agent as counselor extenders, and facilitate accumulation of knowledge. It should also be noted that our methodological approach, and some of our design guidelines, might be useful for the development of virtual agents in different domains, where active listening is important.

Our work moves beyond laboratory studies toward deployable systems by addressing practical concerns (technical usability, diverse user preferences, scalable delivery) while maintaining therapeutic fidelity. Rather than replacing clinicians, our findings suggest that virtual agents as counselors could extend and complement therapeutic reach, particularly valuable for stigmatized population, and given the current mental health workforce shortage.

13 Acknowledgments

We thank the editors and anonymous reviewers for their constructive feedback that enabled us to significantly improved the quality of this article. We also thank participants of the evaluation for helping evaluating and improving our virtual health agents. This research was supported by grant No. 1423260 from the *US National Science Foundation*.

References

- [2] Steven J Ackerman and Mark J Hilsenroth. 2003. A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical psychology review* 23, 1 (Feb. 2003), 1–33.
- [3] Jens Allwood. 2002. Bodily communication dimensions of expression and content. *Multimodality in language and speech systems* 7, 26 (2002), 1–15.
- [4] Nalini Ambady, Jasook Koo, Robert Rosenthal, and Carol H. Winograd. 2002. Physical therapists’ nonverbal communication predicts geriatric patients’ health outcomes. *Psychology and Aging* 17, 3 (2002), 443–452.
- [5] N Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111, 2 (1992), 256–274.
- [6] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi T. Iqbal, Paul N. Bennett, Kori Inkpen, et al. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*. ACM, New York, NY, USA, 1–13.
- [8] Thomas F Babor and Marcus Grant. 1992. *Programme on Substance Abuse : project on identification and management of alcohol-related problems. Report on Phase II, an randomized clinical trial of brief interventions in primary health care*. Technical Report. World Health Organization. <http://apps.who.int/iris/handle/10665/61637>
- [9] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.
- [11] Amy L. Baylor. 2009. Promoting motivation with virtual agents and avatars: Role of visual presence and appearance. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3559–3565.
- [10] Amy L. Baylor and Yanghee Kim. 2004. Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role. In *Intelligent Tutoring Systems: 7th International Conference, ITS 2004 (Lecture Notes in Computer Science, Vol. 3220)*. Springer, Berlin, Heidelberg, 592–603.
- [12] Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud. 2008. A Listening Agent Exhibiting Variable Behaviour. In *Intelligent Virtual Agents: 8th International Conference, IVA 2008 (Lecture Notes in Computer Science, Vol. 5208)*. Springer, Berlin, Heidelberg, 262–269.
- [13] Bridgette M Bewick, Karen Trusler, Michael Barkham, Andrew J Hill, Jane Cahill, and Brendan Mulhern. 2008. The effectiveness of web-based interventions designed to decrease alcohol consumption—a systematic review. *Preventive medicine* 47, 1 (July 2008), 17–26. doi:10.1016/j.yjmed.2008.01.005

- [16] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient education and counseling* 59, 1 (2005), 21–30.
- [14] Timothy Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59, 1 (2005), 21–30.
- [15] Timothy W. Bickmore, Amanda Gruber, and Rosalind Picard. 2005. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling* 59, 1 (2005), 21–30. doi:10.1016/j.pec.2004.09.008
- [17] E. Bordin. 1979. The Generalization of the Psychoanalytic Concept of the Working Alliance. *Psychotherapy: Theory, Research and Practice* 16, 3 (1979), 252–260.
- [18] Hana Boukricha, Ipke Wachsmuth, Maria Nella Carminati, and Pia Knoeferle. 2013. A Computational Model of Empathy: Empirical Evaluation. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Piscataway, NJ, USA, 1–6.
- [19] M. Boustani, S. Lunn, U. Visser, and C. Lisetti. 2021. Development, feasibility, acceptability, and utility of an expressive speech-enabled digital health agent to deliver online, brief motivational interviewing for alcohol misuse: Descriptive study. *Journal of Medical Internet Research* 23, 9 (2021), 1–15.
- [20] Maya Boustani, Stephanie Lunn, Ubbo Visser, and Christine Lisetti. 2021. Development, Feasibility, Acceptability, and Utility of an Expressive Speech-Enabled Digital Health Agent to Deliver Online, Brief Motivational Interviewing for Alcohol Misuse: Descriptive Study. *JMIR mHealth and uHealth* 9, 10 (2021), e30163. doi:10.2196/30163
- [21] Joost Broekens. 2021. Emotion. In *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics, Volume 1: Methods, Behavior, Cognition*. ACM, New York, NY, USA, 349–384.
- [22] Brian L Burke, Hal Arkowitz, and Marisa Menchola. 2003. The efficacy of motivational interviewing: a meta-analysis of controlled clinical trials. *Journal of Consulting and Clinical Psychology* 71, 5 (Oct 2003), 843–861. doi:10.1037/0022-006X.71.5.843
- [23] Valerie A Canady. 2023. More than half of mental health therapists experience burnout. *Mental Health Weekly* 33, 39 (2023), 5–6.
- [24] Kathleen M. Carroll. 2014. Computerized cognitive-behavioral therapy. *Alcohol Research: Current Reviews* 36, 1 (2014), 127–130.
- [25] Kathleen M Carroll. 2014. Lost in translation? Moving contingency management and cognitive behavioral therapy into clinical practice. *Annals of the New York Academy of Sciences* 1327 (Oct 2014), 94–111. doi:10.1111/nyas.12501
- [26] Justine Cassell. 2001. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine* 22, 4 (2001), 67–67.
- [27] Justine Cassell, Hannes Högni Vilhjálmsson, and Timothy Bickmore. 2001. BEAT: The Behavior Expression Animation Toolkit. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*. ACM, New York, NY, USA, 477–486.
- [28] Louis G. Castonguay and Clara E. Hill. 2017. *How and Why Are Some Therapists Better Than Others?: Understanding Therapist Effects*. American Psychological Association, Washington, DC, USA.
- [29] Center for Behavioral Health Statistics and Quality. 2024. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2023 National Survey on Drug Use and Health*. HHS Publication. Substance Abuse and Mental Health Services Administration, Rockville, MD. <https://www.samhsa.gov/data/report/2023-nsduh-annual-national-report>
- [30] Greg Clarke and Bobbi Jo Yarborough. 2013. Evaluating the promise of health IT to enhance/expand the reach of mental health services. *General hospital psychiatry* 35, 4 (2013), 339–344.
- [31] Sarah Clement, Oliver Schuman, Tanya Graham, Francesca Maggioni, Sara Evans-Lacko, Nikita Bezborodovs, Craig Morgan, Nicolas Rüschi, June SL Brown, and Graham Thornicroft. 2015. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological medicine* 45, 1 (2015), 11–27.
- [32] Eric J Cooks, Kyle A Duke, Elizabeth Flood-Grady, Melissa J Vilaro, Rashi Ghosh, Naomi Parker, Palani Te, Thomas J George, Benjamin C Lok, Maribeth Williams, et al. 2022. Can virtual human clinicians help close the gap in colorectal cancer screening for rural adults in the United States? The influence of rural identity on perceptions of virtual human clinicians. *Preventive Medicine Reports* 30 (2022), 102034.
- [33] Paul T. Costa and Robert R. McCrae. 1992. *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL.
- [34] A. Cunningham, J.A., Humphreys, K., & Koski-Jannes. 1999. Providing personalized assessment feedback for problem drinking on the Internet. In *Paper presented at the 33rd annual convention of the Association for the Advancement of Behavior Therapy, Toronto*.
- [35] Rachel G. Curtis, Bethany Bartel, Ty Ferguson, Henry T. Blake, Celine Northcott, Rosa Virgara, and Carol A. Maher. 2021. Improving User Experience of Virtual Health Assistants: Scoping Review. *Journal of Medical Internet Research* 23, 12 (2021), 1–18.
- [36] J. M. Dacy and S. L. Brodsky. 1992. Effects of therapist attire and gender. *Psychotherapy* 29, 3 (1992), 486–490.
- [37] Fred D. Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems* 13, 3 (1989), 319–339. doi:10.2307/249008
- [38] Janis Davis, Michelle Asuncion, Jennifer Rabello, Caitlin Silangcruz, and Elizabeth van Dyk. 2013. A qualitative review of occupational therapists’ listening behaviors and experiences when caring for patients in palliative or hospice care. *OTJR: Occupation, Participation*

- and Health* 33, 1 (2013), 12–20.
- [39] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jonathan Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Pittsburgh, PA, USA, 1061–1068.
- [40] Arie Dijkstra. 2014. The persuasive effects of personalization through name mentioning in a smoking cessation message. *User Modeling and User-Adapted Interaction* 24, 5 (2014), 393–411. doi:10.1007/s11257-014-9147-x
- [41] Katinka Dijkstra, Marcel E Pieterse, and Ad Pruyn. 2008. Stress-reducing effects of indoor plants in the built healthcare environment: The mediating role of perceived attractiveness. *Preventive medicine* 47, 3 (2008), 279–283.
- [42] M Robin DiMatteo. 1979. A Social-Psychological Analysis of Physician-Patient Rapport: Toward a Science of the Art of Medicine. *Journal of Social Issues* 35, 1 (1979), 12–33.
- [43] F. F. Duffy, J. C. West, J. Wilk, W. E. Narrow, D. Hales, J. Thompson, et al. 2002. Mental health practitioners and trainees. In *Mental Health, United States 2002*, R. W. Manderscheid and M. J. Henderson (Eds.). U.S. Department of Health and Human Services, Rockville, MD, 327–368.
- [44] C. Dunn, L Deroo, and FP Rivara. 2001. The use of brief interventions adapted from motivational interviewing across behavioral domains: a systematic review. *Addiction* 96, 12 (2001), 1725–42.
- [45] Chris Dunn, Lisa Deroo, and Frederick P. Rivara. 2001. The use of brief interventions adapted from motivational interviewing across behavioral domains: A systematic review. *Addiction* 96, 12 (2001), 1725–1742.
- [46] Nancy Eisenberg and Janet Strayer. 1990. *Empathy and Its Development*. Cambridge University Press, Cambridge, UK.
- [47] Paul Ekman and Wallace V. Friesen. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto, CA, USA.
- [48] Paul Ekman, Wallace V. Friesen, and Joseph C. Hager. 2002. *Facial Action Coding System: The Manual* (2nd ed.). Research Nexus eBook, Salt Lake City, UT, USA.
- [49] Robert Elliott, Arthur C Bohart, Jeanne C Watson, and Leslie S Greenberg. 2011. Empathy. *Psychotherapy* 48, 1 (2011), 43.
- [50] K M Emmons and S Rollnick. 2001. Motivational interviewing in health care settings. Opportunities and limitations. *American journal of preventive medicine* 20, 1 (Jan. 2001), 68–74.
- [51] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human Computer Studies* 132, June (2019), 138–161.
- [52] Craig Ferguson, Robert Lewis, Chelsey Wilks, and Rosalind W. Picard. 2021. The Guardians: Designing a Game for Long-Term Engagement with Mental Health Therapy. In *2021 IEEE Conference on Games (CoG)*. IEEE, Piscataway, NJ, USA, 1–8.
- [1] Jerome D. Frank. 1973. *Persuasion and healing; a comparative study of psychotherapy* (rev. ed. ed.). Johns Hopkins University Press, Baltimore.
- [53] Carlton Gass. 1984. Therapeutic influence as a function of therapist attire and the seating arrangement in an initial interview. *Journal of Clinical Psychology* 40, 1 (1984), 52–57.
- [54] Shiry Ginosar, Amir Bar, Gefen Kohavi, Caroline Chan, Andrew Owens, and Jitendra Malik. 2019. Learning Individual Styles of Conversational Gesture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, USA, 3497–3506.
- [55] Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine, Chicago, IL, USA.
- [56] Arnold P. Goldstein and Gerald Y. Michaels. 1985. *Empathy: Development, Training, and Consequences*. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.
- [57] Jonathan Gratch. 2025. A Social-Functional View on the Recognition and Analysis of Emotional Expressions. Association for the Advancement of Affective Computing Online Seminar Series. <https://aaac.world/seminars/> Online presentation.
- [58] Jonathan Gratch and Gale Lucas. 2021. Rapport Between Humans and Socially Interactive Agents. *The Handbook on Socially Interactive Agents* 1 (2021), 433–462.
- [59] Jonathan Gratch, Gale Lucas, Aisha King, and Louis-Philippe Morency. 2014. It’s Only a Computer: The Impact of Human-Agent Interaction in Clinical Interviews. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Richland, SC, USA, 85–92.
- [60] Edward T. Hall. 1959. *The Silent Language*. Doubleday, Garden City, NY. 240 pages.
- [61] Edward T. Hall. 1963. A System for the Notation of Proxemic Behavior. *American Anthropologist* 65, 5 (1963), 1003–1026. doi:10.1525/aa.1963.65.5.02a00020
- [62] Edward T. Hall. 1966. *The Hidden Dimension*. Doubleday, Garden City, NY. xii + 193 pages.
- [63] Steven M. Harris and Dean M. Busby. 1998. Therapist physical attractiveness: An unexplored influence on client disclosure. *Journal of Marital and Family Therapy* 24, 2 (1998), 251–257.

- [64] Rex Hartson and Pardha S. Pyla. 2019. *The UX Book: Agile UX Design for a Quality User Experience* (2nd ed.). Morgan Kaufmann, Cambridge, MA, USA.
- [65] Health Resources and Services Administration. 2025. Projecting Health Workforce Supply and Demand. <https://bhwh.hrsa.gov/data-research/projecting-health-workforce-supply-demand> Bureau of Health Workforce, National Center for Health Workforce Analysis. Accessed: 2026-03-07.
- [66] Carolyn J Heckman, Brian L Egleston, and Matthew T Hofmann. 2010. Efficacy of motivational interviewing for smoking cessation: a systematic review and meta-analysis. *Tobacco Control* 19, 5 (Oct 2010), 410–416. doi:10.1136/tc.2009.033175
- [67] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: The almere model. *International Journal of Social Robotics* 2, 4 (2010), 361–375.
- [68] Jeanne R Heitmeyer and Elizabeth B Goldsmith. 1990. Attire, an influence on perceptions of counselors' characteristics. *Perceptual and Motor Skills* 70, 3 (1990), 923–929.
- [69] Monique Hennink and Bonnie N. Kaiser. 2022. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science and Medicine* 292 (2022).
- [70] Stephen G. Henry, Andrea Fuhrel-Forbis, Mary A.M. Rogers, and Susan Eggly. 2012. Association between nonverbal communication during clinical interactions and outcomes: A systematic review and meta-analysis. *Patient Education and Counseling* 86, 3 (2012), 297–315.
- [71] P. Paul Heppner and Steve Pew. 1977. Effects of diplomas, awards, and counselor sex on perceived expertness. *Journal of Counseling Psychology* 24, 2 (1977), 147–149.
- [72] Reid K Hester and Joseph H Miller. 2006. Computer-based tools for diagnosis and treatment of alcohol problems. *Alcohol Research & Health* 29, 1 (2006), 36.
- [73] Reid K. Hester, David D. Squires, and Harold D. Delaney. 2005. The Drinker's Check-Up: 12-Month Outcomes of a Controlled Clinical Trial of a Stand-Alone Software Program for Problem Drinkers. *Journal of Substance Abuse Treatment* 28, 2 (March 2005), 159–169.
- [74] H.D. Hester, R.K., & Delaney. 1997. Behavioral Self-Control Program for Windows: Results of a controlled clinical trial. *Journal of Consulting and Clinical Psychology* 65 (1997), 685–693.
- [75] Darragh Higgins, Donal Egan, Rebecca Fribourg, Benjamin Cowan, and Rachel McDonnell. 2021. Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny?. In *ACM Symposium on applied perception 2021*. ACM, New York, NY, USA, 1–5.
- [76] Clara E. Hill. 1999. *Helping Skills: Facilitating Exploration, Insight, and Action*. American Psychological Association, Washington, DC, USA.
- [77] Mohammadreza Hojat. 2007. *Empathy in Patient Care: Antecedents, Development, Measurement, and Outcomes*. Springer, New York, NY, USA.
- [78] Timothy R Holler. 2006. *The importance of client personality in the prediction of preference for a counseling approach*. Ph.D. Dissertation. The University of Memphis.
- [79] Ghazanfar Hussain, Brian Keegan, and Robert Ross. 2025. Quantifying the role of active listening and reassurance in virtual health coach interactions. *Articles* 244 (2025).
- [80] James M. Jones. 1991. Psychological Models of Race: What Have They Been and What Should They Be? In *Master Lectures in Psychology: Psychological Perspectives on Human Diversity in America*. American Psychological Association, Washington, DC, USA, 3–46.
- [82] Sin-Hwa Kang and Jonathan Gratch. 2014. Exploring users' social responses to computer counseling interviewers' behavior. *Computers in Human Behavior* 34 (2014), 120–130.
- [83] Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. 2017. Audio-driven facial animation by joint end-to-end learning of pose and emotion. *ACM Transactions on Graphics* 36, 4 (2017), 1–12.
- [84] Meghan L. Kelley, Adrian J. Bravo, Erin L. Davies, Katie Witkiewitz, and Ryan Redner. 2022. Treatment of Alcohol Use Problems Among Rural Populations: a Review of Barriers and Considerations for Increasing Access to Quality Care. *Current Addiction Reports* 9 (2022), 511–522. doi:10.1007/s40429-022-00454-3
- [85] Adam. Kendon. 1990. *Conducting interaction : patterns of behavior in focused encounters*. Cambridge University Press, Cambridge ;.
- [86] Alla Keselman, Robert Logan, Catherine Arnott Smith, Gony Leroy, and Qing Zeng-Treitler. 2008. Developing informatics tools and strategies for consumer-centered health communication. *Journal of the American Medical Informatics Association : JAMIA* 15, 4 (2008), 473–83. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2442255&tool=pmcentrez&rendertype=abstract>
- [87] Elly A. Konijn and Johan F. Hoorn. 2005. Some like it bad: Testing a model for perceiving and experiencing fictional characters. *Media Psychology* 7, 2 (2005), 107–144. doi:10.1207/S1532785XMEP0702_1
- [88] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson, and Hannes Vilhjálmsson. 2006. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *Intelligent Virtual Agents: 6th International Working Conference, IVA 2006 (Lecture Notes in Computer Science, Vol. 4133)*. Springer, Berlin, Heidelberg, 205–217.
- [89] Lean L. Kramer, Silke Ter Stal, Bob C. Mulder, Emely De Vet, and Lex Van Velsen. 2020. Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *Journal of Medical Internet Research* 22, 2 (2020), 1–11.

- [90] Clary Krekula, Pirjo Nikander, and Monika Wilińska. 2018. Multiple Marginalizations Based on Age: Gendered Ageism and Beyond. In *Contemporary Perspectives on Ageism*. Springer International Publishing, Cham, Switzerland, 33–50.
- [91] Gary L. Kreps and Linda Neuhauser. 2013. Artificial intelligence and immediacy: Designing health communication to personally engage consumers and providers. *Patient Education and Counseling* 92, 2 (2013), 205–210.
- [92] Taras Kucherenko, Rajmund Nagy, Youngwoo Yoon, Jieyeon Woo, Teodor Nikolov, Mihail Tsakov, and Gustav Eje Henter. 2023. The GENEA Challenge 2023: A Large-Scale Evaluation of Gesture Generation Models in Monadic and Dyadic Settings. In *Proceedings of the 25th International Conference on Multimodal Interaction (ICMI '23)*. ACM, New York, NY, USA, 792–801.
- [93] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction* (2nd ed.). Morgan Kaufmann, San Francisco, CA, USA.
- [94] Cynthia LeRouge, Karen Dickhut, Christine L. Lisetti, Sreejith Sangameswaran, and Thomas Malasanos. 2016. Engaging Adolescents in a Computer-Based Weight Management Program: Avatars and Virtual Coaches Could Help. *Journal of the American Medical Informatics Association* 23, 1 (2016), 19–28.
- [95] Heidi M. Levitt. 2020. *Reporting Qualitative Research in Psychology: How to Meet APA Style Journal Article Reporting Standards*. American Psychological Association, Washington, DC, USA. doi:10.1037/0000179-000
- [96] Patrick Ga Li, Barbara Aquilini, Andrea Davoli, Silvana Grandi, and Chiara Ruini. 2023. The Use of Virtual Reality Interventions to Promote Positive Mental Health: Systematic Literature Review. *JMIR Mental Health* 10 (2023), e44998. doi:10.2196/44998
- [97] Christine Lisetti, Reza Amini, and Ugan Yasavur. 2015. Now all together: overview of virtual health assistants emulating face-to-face health interview experience. *KI-Künstliche Intelligenz* 29, 2 (2015), 161–172.
- [98] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rische. 2013. I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)* 4, 4 (2013), 1–28.
- [99] Christine L. Lisetti and Diane J. Schiano. 2000. Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics & Cognition* 8, 1 (2000), 185–235.
- [100] Christine L. Lisetti, Ugan Yasavur, Claudia De Leon, Reza Amini, Ubbo Visser, and Naphtali Rische. 2012. Building an On-Demand Avatar-Based Health Intervention for Behavior Change. In *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference (FLAIRS-25)*. AAAI Press, Menlo Park, CA, USA, 103–108.
- [101] Judith Lorber. 2018. The Social Construction of Gender. In *Inequality in the 21st Century*, David B. Grusky, Manwai C. Ku, and Szonja Szélényi (Eds.). Routledge, New York, NY, USA, 347–352. <https://doi.org/10.4324/9780429494468>
- [102] Kate Loveys, Gabrielle Sebaratnam, Mark Sagar, and Elizabeth Broadbent. 2020. The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review. *International Journal of Social Robotics* 12, 6 (2020), 1293–1312.
- [103] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37 (2014), 94–100.
- [104] Gale M Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.
- [105] Birgit Lugrin. 2021. Introduction to Socially Interactive Agents. In *The Handbook of Socially Interactive Agents* (1 ed.), Birgit Lugrin, Catherine Pelachaud, and David Traum (Eds.). Vol. 1. Association for Computing Machinery, New York, NY, USA, Chapter 1, 1–18.
- [106] Qiuling Luo, Bruno Rossion, and Milena Dzhelyova. 2019. A robust implicit measure of facial attractiveness discrimination. *Social Cognitive and Affective Neuroscience* 14, 7 (2019), 737–746.
- [107] R. M. Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural Behavior of a Listening Agent. In *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005 (Lecture Notes in Computer Science, Vol. 3661)*. Springer, Berlin, Heidelberg, 25–36.
- [108] Stacy C Marsella and Jonathan Gratch. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10, 1 (2009), 70–90.
- [109] Gabrielle M. Maxwell, Michael W. Cook, and Raymond Burr. 1985. The encoding and decoding of liking from behavioral cues in both auditory and visual channels. *Journal of Nonverbal Behavior* 9, 4 (1985), 239–263.
- [110] Rachel McDonnell, Martin Breidt, and Heinrich H. Bühlhoff. 2012. Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Transactions on Graphics* 31, 4 (2012), 1–11.
- [111] Rachel McDonnell and Bilge Mutlu. 2021. Appearance. In *The Handbook on Socially Interactive Agents, Volume 1: Methods*, Birgit Lugrin, Catherine Pelachaud, and David Traum (Eds.). ACM, New York, NY, USA, 107–146.
- [112] John McKellar, Julia Austin, and Rudolf Moos. 2012. Building the first step: a review of low-intensity interventions for stepped care. *Addiction Science & Clinical Practice* 7, 1 (2012), 26.
- [113] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27 (2001), 415–444. Issue 1.
- [114] Scott W. McQuiggan, Jennifer L. Robison, Robert Phillips, and James C. Lester. 2008. Modeling Parallel and Reactive Empathy in Virtual Agents: An Inductive Approach. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008), Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Richland, SC, USA, 167–174.

- [115] Albert Mehrabian. 1969. Significance of posture and position in the communication of attitude and status relationships. *Psychological Bulletin* 71, 5 (1969), 359–372. doi:10.1037/h0027349
- [116] José Mercado, Ismael Edrein Espinosa-Curiel, and Juan Martínez-Miranda. 2023. Embodied Conversational Agents Providing Motivational Interviewing to Improve Health-Related Behaviors: Scoping Review. *Journal of Medical Internet Research* 25 (2023), e52097. doi:10.2196/52097
- [117] Jörg Merten. 2005. Facial microbehavior and the emotional quality of the therapeutic relationship. *Psychotherapy Research* 15, 3 (2005), 325–333.
- [118] Elizabeth Merwin, Ivora Hinton, Bruce Dembling, and Steven Stern. 2003. Shortages of Rural Mental Health Professionals. *Archives of Psychiatric Nursing* 17, 1 (2003), 42–51.
- [121] William R. Miller and Stephen Rollnick. 2012. *Motivational Interviewing: Helping People Change* (3rd ed.). Guilford Press, New York, NY, USA.
- [122] William R. Miller and Barbara Sanchez. 1994. Motivating Young Adults for Treatment and Lifestyle Change. In *Issues in Alcohol Use and Misuse by Young Adults*. University of Notre Dame Press, Notre Dame, IN, USA, 55–81.
- [123] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The uncanny valley. *IEEE Robotics and Automation Magazine* 19, 2 (2012), 98–100. doi:10.1109/MRA.2012.2192811
- [124] Scott A Myers and Megan F Ferry. 2001. Interpersonal communication motives and nonverbal immediacy behaviors. *Communication Research Reports* 18, 2 (2001), 182–191.
- [125] Jack L Nasar and Ann Sloan Devlin. 2011. Impressions of psychotherapists' offices. *Journal of counseling psychology* 58, 3 (2011), 310–320.
- [126] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology* 27, 10 (1997), 864–876.
- [127] Christina E. Newhill, Elizabeth A. Mulvaney, and Bobby F. Simmons. 2020. Communicating: Verbal Following/Active Listening Skills. In *Skill Development for Generalist Practice: Exercises for Real-World Application*. SAGE Publications, Thousand Oaks, CA, USA, 52–64. <https://us.sagepub.com/en-us/nam/skill-development-for-generalist-practice/book258823>
- [128] Jakob Nielsen. 1994. Heuristic Evaluation. In *Usability Inspection Methods*, Jakob Nielsen and Robert L. Mack (Eds.). John Wiley & Sons, New York, NY, 25–62.
- [129] Jakob Nielsen. 1994 updated 2024. 10 Usability Heuristics for User Interface Design. <https://www.nngroup.com/articles/ten-usability-heuristics/> accessed Jul. 10, 2025.
- [130] Heather L O HL O'Brien, EG Elaine G Toms, By Heather L O'Brien, and EG Elaine G Toms. 2008. What is User Engagement ? A Conceptual Framework for Defining User Engagement with Technology. *Journal of the American Society for Information Science and Technology* 59, 6 (2008), 938–955.
- [131] Magalie Ochs, Radosław Niewiadomski, and C. Pelachaud. 2010. How a virtual agent should smile?. In *Proceedings of the International Conference on Intelligent Virtual Agents*. Springer, Berlin, Heidelberg: Springer Berlin Heidelberg, 427–440.
- [132] Magalie Ochs, David Sadek, and Catherine Pelachaud. 2012. A formal model of emotions for an empathic rational dialog agent. *Autonomous Agents and Multi-Agent Systems* 24, 3 (2012), 410–440.
- [133] Stefan Olafsson, Teresa K O'Leary, and Timothy W Bickmore. 2020. Motivating health behavior change with humorous virtual agents. In *Proceedings of the 20th ACM international conference on intelligent virtual agents*. ACM, New York, NY, USA, 1–8.
- [134] Stanislav Ond'áš, Elisabetta Bevacqua, Jozef Juh'ar, and Peter Demeter. 2010. Towards Influencing of the Conversational Agent Mental State in the Task of Active Listening. In *Development of Multimodal Interfaces: Active Listening and Synchrony*. Springer, Heidelberg, Germany, 95–104.
- [135] World Health Organization et al. 2021. *Mental health atlas 2020*. World Health Organization, Geneva, Switzerland.
- [136] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems* 7, 3 (2017), 1–40. doi:10.1145/2912150
- [137] Lawrence A Palinkas. 2014. Qualitative Methods in Mental Health Services Research. *J Clin Child Adolesc Psychol* 43, 6 (2014), 851–861.
- [138] Dhaval Parmar, Stefan Olafsson, Dina Utami, and Timothy Bickmore. 2018. Looking the part: The effect of attire and setting on perceptions of a virtual health counselor. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents, IVA 2018*. ACM, New York, NY, USA, 301–306.
- [139] Mark Pearson and Helen Wilson. 2012. Soothing spaces and healing places: Is there an ideal counselling room design? *Psychotherapy in Australia* 18, 3 (2012), 46–53.
- [140] Catherine Pelachaud and Isabella Poggi. 2002. Subtleties of facial expressions in embodied agents. *Journal of Visualization and Computer Animation* 13, 5 (2002), 301–312. doi:10.1002/vis.299
- [141] Andre Pereira, Rui Prada, and Ana Paiva. 2014. Improving social presence in human-agent interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, New York, NY, USA, 1449–1458.
- [142] Jairo Perez-Osorio, Eva Wiese, and Agnieszka Wykowska. 2021. Theory of mind and joint attention. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 1*:

- Methods, Behavior, Cognition*. ACM, New York, NY, USA, 311–348.
- [143] David B Portnoy, Lori a J Scott-Sheldon, Blair T Johnson, and Michael P Carey. 2008. Computer-delivered interventions for health promotion and behavioral risk reduction: a meta-analysis of 75 randomized controlled trials, 1988-2007. *Preventive medicine* 47, 1 (July 2008), 3–16. doi:10.1016/j.ypmed.2008.02.014
- [144] Enola K Proctor and Aaron Rosen. 1981. Expectations and preferences for counselor race and their relation to intermediate treatment outcomes. *Journal of Counseling Psychology* 28, 1 (1981), 40.
- [145] Emily Pullen and Carrie Oser. 2014. Barriers to Substance Abuse Treatment in Rural and Urban Communities: A Counselor Perspective. *Substance Use & Misuse* 49, 7 (2014), 891–901. doi:10.3109/10826084.2014.891615
- [146] Ruth Kjørsti Raanaas, Grete Grindal Patil, and Terry Hartig. 2012. Health benefits of a view of nature through the window: A quasi-experimental study of patients in a residential rehabilitation center. *Clinical rehabilitation* 26, 1 (2012), 21–32.
- [147] Hedieh Ranjbartabar, Deborah Richards, Ayse Aysin Bilgin, and Cat Kutay. 2019. First impressions count! The role of the human’s emotional state on rapport established with an empathic versus neutral virtual therapist. *IEEE transactions on affective computing* 12, 3 (2019), 788–800.
- [148] B. Reeves and C. Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. University of Chicago Press, New York, NY.
- [149] Lazlo Ring, Dina Utami, and Timothy Bickmore. 2014. The right agent for the job? The effects of agent visual appearance on task domain. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8637 LNAI (2014), 374–384.
- [150] Albert Rizzo, Scherer Scherer, David DeVault, Jonathan Gratch, Ronald Artstein, Arno Hartholt, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, et al. 2016. Detection and computational analysis of psychological signals using a virtual human interviewing agent. *Journal of Pain Management* 9, 3 (2016), 311–321.
- [151] Kathryn Robertson. 2005. Active listening: more than just paying attention. *Australian family physician* 34, 12 (2005), 1053–1055.
- [152] Scott Robertson, Rob Solomon, Mark Riedl, Theresa Wicklin Gillespie, Toni Chociemski, Viraj Master, and Arun Mohan. 2015. The visual design and implementation of an embodied conversational agent in a shared decision-making context (eCoach). *Lecture Notes in Computer Science* 9192 (2015), 427–437.
- [153] Sérgio Hortas Rodrigues, Samuel Mascarenhas, João Dias, and Ana Paiva. 2015. A process model of empathy for virtual agents. *Interacting with Computers* 27, 4 (2015), 371–391.
- [154] Johanna Roettl and Ralf Terlutter. 2018. The same video game in 2D, 3D or virtual reality – How does technology impact game evaluation and brand placements? *PLoS ONE* 13, 7 (2018), 1–24.
- [155] Carl R. Rogers. 1951. *Client-Centered Therapy: Its Current Practice, Implications and Theory*. Houghton Mifflin, Boston.
- [156] Carl R Rogers. 1975. Empathic: An unappreciated way of being. *The counseling psychologist* 5, 2 (1975), 2–10.
- [157] Astrid Rosenthal-von der Pütten and Anna M. H. Abrams. 2021. Empirical Methods in the Social Science for Researching Socially Interactive Agents. *The Handbook on Socially Interactive Agents* 1 (2021), 19–76.
- [158] Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations. *Artificial Intelligence in Cybersecurity (AICS)* 2563 (2019), 104–115.
- [159] Sune Rubak, Anneli Sandbaek, Torsten Lauritzen, and Bo Christensen. 2005. Motivational interviewing: a systematic review and meta-analysis. *British Journal of General Practice* 55, 513 (Apr 2005), 305–312.
- [160] Catharina Rudschies and Ingrid Schneider. 2024. Ethical, legal, and social implications (ELSI) of virtual agents and virtual reality in healthcare. *Social Science & Medicine* 340 (2024), 116483.
- [161] Mary S. Ryan. 1968. *Clothing: A Study in Human Behavior*. Holt, Rinehart and Winston, New York, NY, USA.
- [162] Jennifer Sabourin, Bradford Mott, and James Lester. 2011. Computational Models of Affect and Empathy for Pedagogical Virtual Agents. *Standards in Emotion Modeling* (2011), 1–14.
- [163] Heidi Salonen, Marjaana Lahtinen, Sanna Lappalainen, Luke D Knibbs, Lidia Morawska, Kari Reijula, et al. 2014. The impact of windows, daylight and views of nature on health and wellbeing in healthcare facilities. In *The International Sustainable Built Environment Conference 2014*. Helsinki, Finland.
- [164] M. L. Schoeneberger, C. G. Leukefeld, M. L. Hiller, and T. Godlaski. 2006. Substance abuse among rural and very rural drug users at treatment entry. *Am J Drug Alcohol Abuse* 32, 1 (2006), 87–110. doi:10.1080/00952990500328687 Schoeneberger, Marlies L Leukefeld, Carl G Hiller, Matthew L Godlaski, Ted 6 UR1 TI 11613-01-1/TI/CSAT SAMHSA HHS/United States Comparative Study Journal Article Research Support, U.S. Gov’t, P.H.S. England Am J Drug Alcohol Abuse. 2006;32(1):87-110. doi: 10.1080/00952990500328687..
- [165] Mark R Scholten, Saskia M Kelders, and Julia EWC Van Gemert-Pijnen. 2017. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of medical Internet research* 19, 11 (2017), e7351.
- [166] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wollmer. 2012. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3, 2 (2012), 165–183. doi:10.1109/T-AFFC.2011.34

- [167] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary Mckeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar, and Martin Wollmer. 2012. Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing* 3, 2 (2012), 165–183.
- [168] Christy K Scott, Christine E Grella, Michael L Dennis, and Lora Nicholson. 2018. Linking Individuals with Substance Use Disorders (SUDs) in Primary Care to SUD Treatment: the Recovery Management Checkups-Primary Care (RMC-PC) Pilot Study. *Journal of Behavioral Health Services & Research* 45, 2 (Apr 2018), 160–173. doi:10.1007/s11414-017-9576-5
- [169] Mike Seymour, Lingyao Yuan, Alan R. Dennis, and Kai Riemer. 2019. Crossing the uncanny valley? Understanding affinity, trustworthiness, and preference for more realistic virtual humans in immersive environments. *Proceedings of the Annual Hawaii International Conference on System Sciences* 2019-January (2019), 1748–1758.
- [170] Jocelyn Shen, Kimaya Lecamwasam, Hae Won Park, Cynthia Breazeal, and Rosalind Picard. 2023. Designing Conversational Agents for Emotional Self-Awareness. In *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. IEEE, Cambridge, MA, USA, 1–4.
- [171] Sarah L Sinclair et al. 2020. Client and Therapists’ Subjective Understanding of an Ideal Therapy Room: A Divergent Reflection of Experience. *The European Journal of Counselling Psychology* 9, 1 (2020), 91–127. doi:10.5964/ejcop.v9i1.201
- [172] T. Sinclair. 2021. What’s in a therapy room?—A mixed-methods study exploring clients’ and therapists’ views and experiences of the physical environment of the therapy room. *Counselling and Psychotherapy Research* 21, 1 (2021), 118–129.
- [173] H. A. Skinner. 1994. Computerized Lifestyle Assessment.
- [174] Agnieszka Sorokowska, Piotr Sorokowski, Peter Hilpert, Katarzyna Cantarero, Tomasz Frackowiak, Khodabakhsh Ahmadi, Ahmad M. Alghraibeh, Richmond Aryeetey, Anna Bertoni, Karim Bettache, et al. 2017. Preferred Interpersonal Distances: A Global Comparison. *Journal of Cross-Cultural Psychology* 48, 4 (2017), 577–592. doi:10.1177/0022022117698039
- [175] Ian Steenstra, Farnaz Nouraei, Mehdi Arjmand, and Timothy Bickmore. 2024. Virtual agents for alcohol use counseling: Exploring IIm-powered motivational interviewing. In *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*. ACM, New York, NY, USA, 1–10.
- [176] Carolin Straßmann, Nicole C. Krämer, Hendrik Buschmeier, and Stefan Kopp. 2020. Age-Related Differences in the Evaluation of a Virtual Health Agent’s Appearance and Embodiment in a Health-Related Interaction: Experimental Lab Study. *Journal of Medical Internet Research* 22, 4 (2020), e13726. doi:10.2196/13726
- [178] Substance Abuse and Mental Health Services Administration. 2015. *Racial/Ethnic Differences in Mental Health Service Use among Adults*. Technical Report HHS Publication No. SMA-15-4906. Substance Abuse and Mental Health Services Administration, Rockville, MD. <https://www.integration.samhsa.gov/MHServicesUseAmongAdults.pdf> Retrieved March 27, 2019.
- [179] Colm Sweeney, Courtney Potts, Edel Ennis, Raymond Bond, Maurice D. Mulvenna, Siobhan O’Neill, Martin Malcolm, Lauri Kuosmanen, Catrine Kostenius, Alex Vakaloudis, Gavin McConvey, Robin Turkington, David Hanna, Heidi Nieminen, Anna Kaisa Vartiainen, Alison Robertson, and Michael F. McTear. 2021. Can Chatbots Help Support a Person’s Mental Health? Perceptions and Views from Mental Healthcare Professionals and Experts. *ACM Transactions on Computing for Healthcare* 2, 3 (2021), 1–15.
- [180] Silke ter Stal, Lean Leonie, Monique Tabak, Den Akker, and Hermie Hermens. 2020. Design Features of Embodied Conversational Agents in eHealth: a Literature Review. *Journal of Human Computer Studies* 138, January (2020), 102409.
- [181] Ilaria Torre, Simon Holk, Emma Carrigan, Iolanda Leite, Rachel McDonnell, and Naomi Harte. 2021. Dimensional Perception of a ‘Smiling McGurk Effect’. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Nara, Japan, 1–8.
- [182] Tom Townsend. 2011. Patient privacy and mental health care in the rural setting. *AMA Journal of Ethics* 13, 5 (2011), 282–286.
- [183] David Traum. 2008. Talking to virtual humans: Dialogue models and methodologies for embodied conversational agents. In *Modeling communication with robots and virtual humans: second zIF research group international workshop on embodied communication in humans and machines, Bielefeld, Germany, April 5-8, 2006, revised selected papers (Lecture Notes in Computer Science, Vol. 4930)*. Springer, Berlin, Heidelberg, 296–309.
- [184] Roger S. Ulrich. 1984. View through a window may influence recovery from surgery. *Science* 224, 4647 (1984), 420–421.
- [185] U.S. Census Bureau. 2024. 2019-2023 American Community Survey 5-Year Estimates: Race and Hispanic Origin. <https://data.census.gov/Table DP05: ACS Demographic and Housing Estimates>, Retrieved July 11, 2025.
- [186] Luis A Valdez, David O Garcia, John Ruiz, Eyal Oren, and Scott Carvajal. 2018. Exploring structural, sociocultural, and individual barriers to alcohol abuse treatment among Hispanic men. *American Journal of Men’s Health* 12, 6 (Nov 2018), 1948–1957. doi:10.1177/1557988318790882
- [187] Henriette C van Vugt, Johan F Hoorn, and Elly A Konijn. 2009. Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds* 20, 2-3 (2009), 195–204.
- [188] Viswanath Venkatesh, Michael G Morris, Gordon B Davis, and Fred D Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS quarterly* 27, 3 (2003), 425–478. doi:10.2307/30036540
- [189] ML Vernon. 2010. A review of computer-based alcohol problem services designed for the general public. *Journal of substance abuse treatment* 38, 3 (2010), 203–211. doi:10.1016/j.jsat.2009.11.001.A

- [190] Maria Wamsley, Jason Satterfield, Aimee Curtis, Lena Lundgren, and Derek Satre. 2018. Alcohol and Drug Screening, Brief Intervention, and Referral to Treatment (SBIRT) Training and Implementation: Perspectives from 4 Health Professions. *Journal of Addiction Medicine* 12, 4 (2018), 262–272. doi:10.1097/ADM.0000000000000410
- [191] Harry Weger, Gina R. Castle, and Melissa C. Emmett. 2010. Active listening in peer interviews: The influence of message paraphrasing on perceptions of listening skill. *International Journal of Listening* 24, 1 (2010), 34–49. doi:10.1080/10904010903466311
- [192] Harry Weger Jr, Gina Castle Bell, Elizabeth M Minei, and Melissa C Robinson. 2014. The relative effectiveness of active listening in initial interactions. *International Journal of Listening* 28, 1 (2014), 13–31.
- [193] Whole Building Design Guide. 2023. Therapeutic Environments. <https://www.wbdg.org/resources/therapeutic-environments> Accessed July 2025.
- [194] Danyell Wilson-Howard, Melissa J. Vilaro, Jordan M. Neil, Eric J. Cooks, Lauren N. Griffin, Taylor T. Ashley, Fatemeh Tavassoli, Mohan S. Zalake, Benjamin C. Lok, Folakemi T. Odedina, Francois Modave, Peter J. Carek, Thomas J. George, and Janice L. Krieger. 2021. Development of a Credible Virtual Clinician Promoting Colorectal Cancer Screening via Telehealth Apps for and by Black Men: Qualitative Study. *JMIR Formative Research* 5, 12 (2021), e28709.
- [195] Jiyeon Woo, Kazuhiro Shidara, Catherine Achard, Hiroki Tanaka, Satoshi Nakamura, and Catherine Pelachaud. 2024. Adaptive virtual agent: Design and evaluation for real-time human-agent interaction. *International Journal of Human-Computer Studies* 190 (2024), 103321.
- [196] World Health Organization. 2025. Over a Billion People Living with Mental Health Conditions: Services Require Urgent Scale-Up. <https://www.who.int/news/item/02-09-2025-over-a-billion-people-living-with-mental-health-conditions-services-require-urgent-scale-up> Accessed: 2026-03-07.
- [197] Özge Nilay Yalçın. 2020. Empathy framework for embodied conversational agents. *Cognitive Systems Research* 59 (2020), 123–132.
- [198] U. Yasavur, C. L. Lisetti, and N. D. Rishe. 2014. Let’s Talk! Speaking Virtual Counselor Offers You a Brief Intervention. *Journal on Multimodal User Interfaces* 8, 4 (2014), 381–398. doi:10.1007/s12193-014-0169-y
- [199] So Yeon Yoon, James Laffey, and Hyunjoo Oh. 2008. Understanding usability and user experience of web-based 3D graphics technology. *International Journal of Human-Computer Interaction* 24, 3 (2008), 288–306.
- [200] Youngwoo Yoon, Bok Cha, Joo-Haeng Lee, Minsu Jang, Jaeyeon Lee, Jaehong Kim, and Geehyuk Lee. 2020. Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 1–16.

A Code book

Table 17. Definition of theme codes and subcodes, frequency counts of users' comments related to theme (n), and inter-rater reliability (κ)

Theme Codes/Subcodes	Definition - Comments that refer to...	n	κ
Agent acceptability			
General user experience	personal feelings or emotional response stemming from the experience with the initial agents selection menu and/or the interaction with the agents.	12	0.79
Personal qualities	perceived personal qualities of the agent, such as friendliness, coldness, competency, etc.	11	0.90
Appearance	visual appearance, physical features, or associated visual components of the agent such as clothing, glasses, and other accessories.	24	0.88
Gestures/Non-Verbals	gestures, body language, and non-verbal communication of the agent.	27	0.90
Voice/verbal tone	voice and verbal features of the agent, e.g., verbal expression, intonation, cadence.	18	0.90
Voice speed	speed of the voice of the agent.	23	1.00
Agent diversity			
Ethnicity/race	perceived ethnicity / race of the agent.	14	0.59
Age	perceived age of the agent.	11	0.75
Gender	perceived gender of the agent.	18	0.80
Environment			
General user experience	reported feelings about the virtual background environment around the agent.	3	1.00
Realism	level of realism of the environment.	5	0.66
Preferences/view	preferences for office environment and views outside the office windows.	17	0.79

B State of the Art on virtual agents as Counselors

Table 18. Overview of Healthcare-Related Agents and Guidelines Developed

Reference [Author(s), (Year), Citation]	System	Purpose/ Research Questions/ Hypotheses	Research Design	Participants	Quantitative Only: (Y/N and if Y, Methods Employed)	Mixed Methods: (Y/N and if Y, Methods Employed)	Key Findings	Guidelines Suggested
Bickmore et al., (2005), [14]	A relational agent employs emotional and relational behaviors within an exercise adoption program.	Pilot study to evaluate the effectiveness of having the agent use emotional and relational communication strategies	Randomized, parallel-group, six-week trial; three treatment groups in the study: Control, Non-Relational, Relational	Generally healthy adults who were interested in becoming more physically active, but were not yet maintaining the recommended 30 min, per day of moderate activity	N	Y, Quantitative: Scores on the Working Alliance Inventory (WAI); Qualitative: Interviews	Patients generally receptive to the technology and responded in ways analogous to emotional and relational communication with human health providers	Computer systems interacting with patients, especially those with dialogue or in long-term, repeated interactions, should consider explicitly designing in emotional and relational communication behavior
Lucas et al., (2014), [103]	A kiosk-based human for anonymous, rapport-building, self-administered screening interviews.	Predicted that, while interacting with a virtual agent (VH)-interviewer, holding the belief that the VH is run automatically by the computer – without oversight by a human being – will increase willingness to disclose. Specifically, being told that the VH is automated should reduce psychological barriers to disclosure (fear of self-disclosure, impression management) and increase open, disclosive behavior in the interview	A, 2 (frame: computer vs. human) x 2 (method: tele-operated vs. AI) design; All participants interacted with a rapport-building VH; The VH was either tele-operated by humans or fully-automated. In the computer frame, participants were led to believe that their responses were not being observed. In the human frame condition, participants were led to believe that their responses were being observed.	239 participants (149 males, 90 females) were recruited via Craigslist. All met requirements (i.e., age between 18 and 65 and adequate eyesight) were accepted.	N	Y, Quantitative: Questionnaires to measure fear of self-disclosure, a modified version of the Brief Fear of Negative Evaluation Scale – Straightforward Items; Observer rated willingness to disclose based on responses; Means and standard errors; 2 (frame: computer vs. human) x 2 (method: tele-operated vs. AI) ANOVAs on each measure of willingness to disclose; Qualitative: Answers to questions and informal anecdotal quotes provided	Participants who were told that the virtual human interviewer was fully automated reported significantly lower fear of self-disclosure and engaged in less impression management; Virtual humans increase willingness to disclose; Compared to participants who were told that the virtual agent was tele-operated by a human, those who were told that the virtual agent interviewer was fully automated allowed themselves to display more intense expressions of sadness; Framing the interviewer as computer lowers evaluation fears and impression management	VH-interviewers can develop rapport and administer more complex questionnaires without a human interviewer – and therefore also without feeling as though responses are being judged
Parmar et al., (2018), Parmar 2018a	3D medical office and agent, web-based scene with a 5-minute interactive conversation, where agent provided information about a serious healthcare topic. The agent used synthesized speech accompanied by automatically generated nonverbal behavior using BEAT.	H1: An agent wearing role-appropriate attire will be rated higher on trustworthiness and credibility, and will be more persuasive, compared to the same agent wearing casual attire; H2: An agent in a role-appropriate virtual setting will be rated higher on trustworthiness and credibility, and will be more persuasive, compared to the same agent in an empty room	Study 1: Within-subjects design, included socio-demographic and health literacy questionnaire to elicit user feedback on attire and setting options, where VH and environment evaluated on professionalism, trustworthiness, credibility, persuasiveness. Study 2: 2x2 factorial (Attire vs. Setting) between-subjects study.	Study 1: 43 participants (23 Male, 20 Female) between the ages of 16-62, of which 28 successfully completed the questionnaire; 308 participants (133 Male, 175 Female) aged 18-76	Y, Paired samples Wilcoxon test for professionalism of attire and setting; Counts of objects in a typical doctor's office; Perceived Professionalism; Shapiro-Wilk test for normality; two-way independent measures ANOVA	N	Agent rated as more professional when in a white coat and in the full office setting; No interaction between attire and setting for credibility of information; Agent in white coat was seen as more trustworthy; Reassuring, caring, friendly, comfortable to talk to, and more appropriate for the job; Wanted to continue working with white coat agent more than casually dressed agent; Impact of setting, full office more realistic than empty room; Participants more committed to choosing a healthcare proxy for themselves in professional attire	Participants favor physicians wearing professional attire and white coats and could lead to increased persuasion to choose a healthcare proxy; Felt more like a face-to-face conversation with white coat agent. No effects on setting of empty room relative to medical office. Concluded designers should focus more on designing role-appropriate attire when providing information on a serious healthcare topic and when motivating them to make decisions.

Reference [Author(s), (Year), Citation]	System	Purpose/ Research Questions/ Hypotheses	Research Design	Participants	Quantitative Only: (Y/N and if Y, Methods Employed)	Mixed Methods: (Y/N and if Y, Methods Employed)	Key Findings	Guidelines Suggested
Olafsson et al., (2020), [133]	Standalone application for Windows run on screen tablet. System has a 3D animated character that converses with users using synthetic speech output and multiple-choice menu inputs for user responses, with allowed user responses updated at each turn of the conversation. The system generates synchronized non-verbal conversational behavior for the agent using BEAT, including hand gestures, head nods, eyebrow raises, posture shifts, and gaze behavior. Agent utterances and menu options are generated using template-based text generation.	H1: Participants will accept an VH from MI to promote physical activity and eating the recommended amount of fruits and vegetables. H2: Participants will be more accepting of the VH that uses stage and topic-appropriate affiliative humor than the one that does not. H3: The conversation with the VH that uses stage- and topic-appropriate affiliative humor will have a greater positive impact on participants' attitudes towards target behavior in the VH than with the does not.	Within-subjects design with two conditions. 1. The VH tells four jokes during counseling session relevant to the topic and whether they regularly engaged in the target behavior (e.g., exercise regularly or eat a healthy diet) or not (HUMOUR). 2. No humor in a similar counseling session (NoHUMOUR). Treatment order, VH appearance, and health topic (exercise or nutrition) were counterbalanced.	15 research participants were recruited using an online advertisement, were required to be 21 years or older, able to speak and read English, and self-report as being in one of the first three stages of change with respect to the target behaviors. Participants were 36.73 (12.65) years old, 40% female, 73% white, 91% single, and all had either some college experience or a college degree	N	Y, Quantitative: Prior to interacting with the agents, collected: Socio-demographics, Situational Humor Response Questionnaire, Humor Styles Questionnaire, Sense of Humor Questionnaire, Need for Cognition Scale, and the Big-Five personality traits questionnaire. For (exercise and nutrition) before the interaction and after: Decisional Balance, Self-Efficacy, Motivation, Confidence, and Stage of Change. The final set of measures were collected following each of the agent conversations: General agent ratings and funniness, Interpersonal Trust, and Bond-subscale of the Working Alliance Inventory. Following the second agent conversation, participants could choose which agent they wanted to talk to; ANOVA run; Structured Interviews coupled with grounded theory and selective coding	Agent in Humour condition rated funnier than no humour condition. Participants significantly greater change in motivation to engage in target behavior following humour agent condition. Also, requested to interact with this condition for an additional conversation; Based on qualitative feedback, participants described the humorous agent as funny, found the humor to be more subtle and light-hearted and came away from the interaction with a sense of validation about their own health concerns	Overall, participants accepted and endorsed the use of humor during the health counseling conversations with an agent and highlighted the benefits of affiliative humor and most focused on how humor made the intervention more engaging. This was reflective of benefits similar to patient-provider communication. Humor also made them more likely to re-initiate the conversation about the humorous agent, indicating that humor may be a useful tool for increasing user engagement. However, participants who reported health-related difficulties, where self-identified lack of exercise and poor nutrition were directly impacting their quality-of-life, perceived the humorous agent as less kind. The authors suggest more research into the effects of humor in designs of virtual agents.
Ferguson et al., (2021), [52]	A free-to-play mobile game that incentivizes players to regularly complete healthy real-world tasks in exchange for in-game reward	Use game for behavioral activation (BA) - an evidence-based psychotherapy that helps individuals to reduce depressive symptoms, through a combination of features including tracking perceived correlations between mood and activities, engaging in adaptive activities, and receiving psycho-education; Sought to explore engagement and effectiveness.	Character and narrative design-focus group; Real-World Dataset	14 individuals in a focus group; Anonymous user-level game statistics of 7,782 users over 54,461 interaction days	N	Y, Quantitative: Explored day-n (since installation) retention relative to mental health apps and mobile games, as well as for game play and task completion; Also investigated engagement with specific features; Qualitative: Focus groups involving 14 individuals were conducted early in the design process	Game improved long-term player engagement relative to digital mental health interventions; A broad range of activities were performed across the main categories and on average they helped users to feel at least a little better; Fitness activities were the most effective	None, although mentioned that rewards mechanism serves as incentive for players

Reference [Author(s), (Year), Citation]	Purpose/ Research Questions/ Hypotheses	Research Design	Participants	Quantitative Only: (V/N and if Y, Methods Employed)	Mixed Methods: (V/N and if Y, Methods Employed)	Key Findings	Guidelines Suggested
Shen et al., (2023), [170]	Develop emotional awareness and well-being through interactions with conversational artificially intelligent agents (AIs)	Pilot study, interacting with the agent daily over 3 days and recorded their emotional self-awareness and mental well-being. Experimental condition and control group	6 graduate and undergraduate students age 18 or older - 4 in experimental and two in control	Y, Quantitative: Pre-study questionnaire on demographics, positive and negative affect (Affect Balance Scale), personality traits (Big Five Personality Test), short-term current mood (Brief Mood Introspection Scale), long-term perception of stress (Perceived Stress Scale) and alexithymia, difficulties in expressing emotions, (Toronto Alexithymia Scale); Also, success of the therapeutic alliance between the participant and the system via the Working Alliance Inventory; Qualitative: Interviews	Presented an affinity diagram to identify seven major design insights (Privacy and control; agent's persona; agent's social-emotional skills, personalization, interaction modality, interventions, self-awareness); Agent's skills were off-putting to all the participants, requested responses more personalized	Users felt a need for control over their data and suggested having cues for when the app is recording them and ways to delete or manage what they told the agent; Despite privacy concerns, the interaction modality of speech was very important for the specific application of improving emotional awareness; Persona important for self-disclosure	
Kang and Kang, (2023), [?]	Chat-bot with an anthropomorphic interface. Included two gender options and the agent's personality or introvert or extrovert manipulated through expression elements (e.g., hejjes, back-channel behaviors)	Experiments with a 2 (personality of the chatbot: introvert/extrovert) x 2 (gender of the chatbot: male/female) x 2 (visual interface cue: yes/no) x 2 (gender of the participant: male/female) between-subjects design	109 participants ranging from 20 to 36 years old from online and offline bulletin boards serving university students in Korea	Y, Self-disclosure measured (using the revised self-disclosure scale for control of disclosure) and depth and honesty of companionship; multi-way analysis of variance (ANOVA) to test hypotheses for effects of the chatbot's gender, personality, and visual interface cue and participant's gender on self-disclosure and companionship; Ran factor analysis, means, standard deviations, t-tests, independent sample and interaction analyses	Y, Self-disclosure measured (using the revised self-disclosure scale for control of disclosure) and depth and honesty of companionship; multi-way analysis of variance (ANOVA) to test hypotheses for effects of the chatbot's gender, personality, and visual interface cue and participant's gender on self-disclosure and companionship; Ran factor analysis, means, standard deviations, t-tests, independent sample and interaction analyses	Chatbots with an anthropomorphized visual interface cue had a negative effect on the self-disclosure and companionship of the participant. When a chatbot of the opposite gender to the participant was provided, the control of depth of disclosure increased. In addition, the results indicated that differences based on the gender of the chatbot were greater with the visual interface cue than that without the interface cue	Self-disclosure and companionship vary depending on the chatbot's anthropomorphic and user characteristics - suggesting that chatbots should be customized to each user

C I-PEFiC Original Model

The I-PEFiC model we applied in this research is shown in Figure 8.

Received 02 May 2023; revised 12 June 2023; accepted 5 July 2023

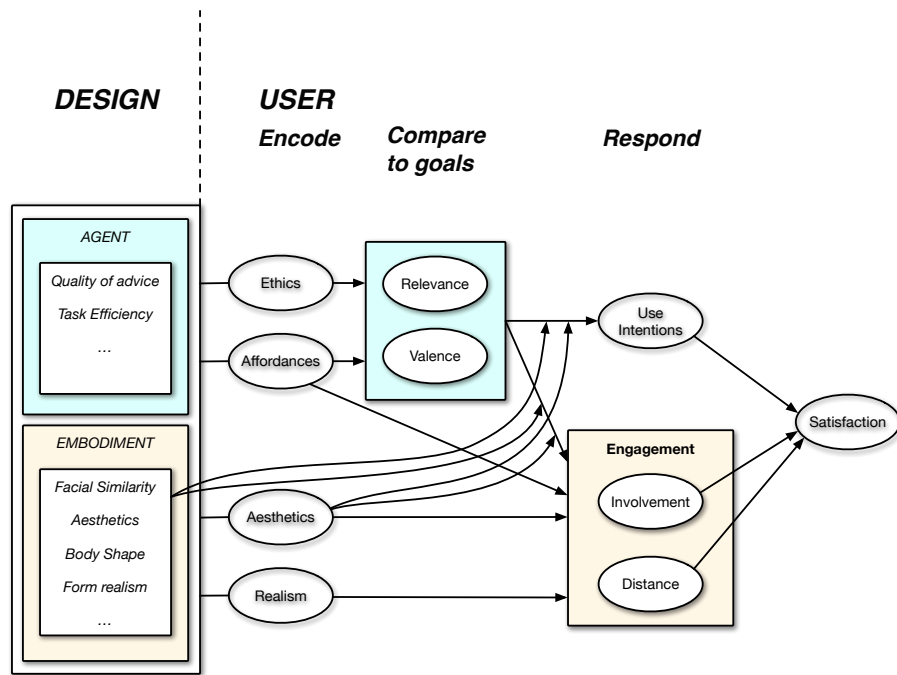


Fig. 8. Interactively Perceiving and Experiencing Fictional Characters (I-PEFiC) model [187].